



Modelo de diálogo aplicado en la terapia de reminiscencia de personas con demencia

por

Víctor Manuel Morales de Jesús

Tesis presentada en cumplimiento parcial de los requisitos
para la obtención del grado de:

Doctor en Ingeniería del Lenguaje y del Conocimiento

Benemérita Universidad Autónoma de Puebla,
Facultad de Ciencias de la Computación

Directores de tesis:

Dra. María Josefa Somodevilla García
Dra. Helena Montserrat Gómez Adorno

Puebla, México, Noviembre 2021.

Dedico esta tesis a mi familia

Resumen

La terapia de reminiscencia es una intervención no farmacológica que ayuda a mitigar los estados psicológicos y emocionales inestables de los pacientes con la enfermedad de Alzheimer, en la que se evocan experiencias del pasado a través de conversaciones entre los pacientes y sus cuidadores, estimulando la memoria episódica autobiográfica. Es altamente recomendable que las personas con Alzheimer reciban regularmente este tipo de terapia. Sin embargo, debido a la carga de trabajo de los cuidadores, es casi imposible que brinden la terapia de forma constante. Por lo que en este trabajo de investigación, proponemos el desarrollo de un sistema conversacional que puede ser utilizado como herramienta para proporcionar terapia de reminiscencia a personas con demencia de tipo Alzheimer en etapa inicial. El sistema tiene la capacidad de personalizar la terapia según la información de los pacientes relacionada con sus preferencias, historia y estilo de vida. Una evaluación realizada por once evaluadores relacionados con el cuidado de personas con Alzheimer como cuidadores (n=9), médico geriatra (n=1) y asistente del centro de cuidado (n=1) nos permitió determinar que el sistema es capaz de llevar a cabo una terapia de reminiscencia de acuerdo con la información del paciente de forma satisfactoria.

Abstract

Reminiscence therapy is a non-pharmacological intervention that helps mitigate unstable psychological and emotional states in patients with Alzheimer's disease, where past experiences are evoked through conversations between the patients and their caregivers, stimulating autobiographical episodic memory. It is highly recommended that people with Alzheimer regularly receive this type of therapy. However, due to the overload of caregivers, it is almost impossible for them to provide therapy on a regular basis. In this paper, we describe the development of a conversational system that can be used as a tool to provide reminiscence therapy to people with Alzheimer's disease. The system has the ability to personalize the therapy according to the patients information related to their preferences, life history and lifestyle. An evaluation conducted with eleven evaluators related to people with Alzheimer's disease care such as caregiver (n=9), geriatric doctor (n=1), and a care center assistant shows that the system is capable of carrying out a reminiscence therapy according to the patient information in a successful manner.

Agradecimientos

Quiero agradecer especialmente a mis asesoras de tesis, Dra. Josefa Somodevilla y Dra. Helena Gómez por su inconmensurable apoyo durante todo este proceso. Sin duda alguna, sin todo su apoyo, paciencia y aliento esto no hubiera sido posible.

También agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por otorgarme la beca No. 626812 la cual me brindó la posibilidad de culminar mis estudios de posgrado.

Índice general

Resumen	II
Abstract	III
Agradecimientos	IV
Índice de Figuras	VII
Índice de Tablas	VIII
1. Introducción	1
1.1. Problema de investigación	3
1.2. Objetivos de la investigación	3
1.2.1. Objetivo general	3
1.2.2. Objetivos específicos	4
1.3. Preguntas de investigación	4
1.4. Organización de la tesis	4
2. Marco teórico	6
2.1. Terapia de reminiscencia	6
2.2. Sistema de diálogo	7
2.2.1. Reconocimiento automático del habla	9
2.2.2. Comprensión del lenguaje natural	9
2.2.3. Gestor de diálogo	10
2.2.3.1. Estrategias de control de diálogo	11
2.2.3.2. Modelos de control del diálogo	12
2.2.4. Generación de lenguaje natural	12
2.2.5. Conversión de texto a habla	13
2.3. Clasificación de los sistemas de diálogo	13
2.3.1. Sistemas de diálogo enfocados a tarea	14
2.3.2. Sistemas de diálogo de pregunta-respuesta	14
2.3.3. Sistemas de diálogo conversacionales	16
2.4. Métricas de evaluación de sistemas de diálogo conversacionales	17
2.5. AIML	19

2.6. Conclusión del capítulo	20
3. Estado del arte	21
3.1. Trabajos iniciales en el modelado del diálogo	21
3.2. Sistemas de diálogo en el dominio de la demencia	23
3.2.1. Sistemas enfocados al diagnóstico	23
3.2.2. Sistemas enfocados al apoyo cognitivo	25
3.2.3. Sistemas enfocados a terapia	26
3.2.3.1. Trabajos en el idioma español	28
3.3. Conclusión del capítulo	29
4. Metodología	30
4.1. Prototipo del Sistema Conversacional	30
4.1.1. Módulo ASR	32
4.1.2. Módulo de Preprocesamiento de Texto	34
4.1.3. Gestor de Diálogo	34
4.1.3.1. Base de datos de pacientes	35
4.1.3.2. Modelo de diálogo	38
4.1.3.3. Personalización del diálogo	41
4.1.4. Módulo TTS	42
4.2. Metodología para llevar acabo la terapia de reminiscencia por el sistema conversacional	43
4.2.1. Características de los participantes	43
4.2.2. Diseño de las intervenciones	45
5. Experimentos y resultados	46
5.1. Comparación evaluativa de sistemas ASR	46
5.1.1. Conjunto de datos	47
5.1.2. Experimentos	48
5.1.3. Resultados	50
5.2. Evaluación del sistema conversacional	51
5.2.1. Configuración de la evaluación	51
5.2.2. Resultados	53
6. Conclusiones	57
6.1. Conclusiones	57
6.2. Limitaciones	58
6.3. Aportaciones	59
6.4. Trabajo futuro	60

Índice de figuras

2.1. Arquitectura clásica de un sistema de diálogo	8
2.2. Jerarquía de etiquetas en objetos AIML	20
4.1. Arquitectura del prototipo del sistema conversacional	31
4.2. Implementación del Módulo ASR	32
4.3. Recolección de información de pacientes	36
4.4. Formulario para el registro de información personal	36
4.5. Formulario para el registro de relaciones familiares	37
4.7. Modelo de la base de datos Pacientes	37
4.6. Formulario para el registro de historia y estilo de vida	38
4.8. Ejemplo de la definición en AIML del contexto <i>Conversación inicial</i>	40
4.9. Reducción simbólica aplicada en el contexto de “cantante”	40
4.10. Plantillas personalizadas con información del paciente	41
4.11. Implementación del Módulo TTS	43
5.1. Comparación de los valores WER obtenidos	50
5.2. Ejemplo de una conversación donde el evaluador tiene que seleccionar la respuesta más apropiada de acuerdo al contexto de diálogo .	52
5.3. Valores R_{voted} obtenidos por cada respuesta en el conjunto de diálogos evaluados	54
5.4. Correlación entre los valores $R_{voted}(res_sys_t, context_t)$ y R_{max} . Una pequeña cantidad de sombra fue añadida a los valores R_{voted} para visualizar la mayoría de los casos	55
5.5. Porcentaje de coincidencia entre las respuestas candidatas y la respuesta más votada.	56

Índice de tablas

3.1. Estudios centrados en el desarrollo de sistemas interactivos en el dominio de la demencia	24
4.1. Contextos de diálogo definidos. Los contextos descritos son sobre los cuales el sistema tiene la capacidad de llevar a cabo la terapia. .	39
4.2. Descripción de síntomas en PcA en etapa inicial	44
5.1. Descripción de archivos de audio	48
5.2. Valores WER obtenidos para cada archivo de audio	49

Capítulo 1

Introducción

La Organización Mundial de la Salud (OMS) describe a la demencia como un síndrome; es decir, un conjunto de síntomas usualmente de naturaleza crónica y progresiva, causado por una variedad de enfermedades cerebrales que afectan la memoria, el pensamiento, el comportamiento y la capacidad para realizar las actividades diarias (World Health Organization, 2020).

La demencia suele afectar principalmente a personas a mayores de 65 años (adultos mayores). El Alzheimer es el tipo de demencia más común, abarcando entre el 60 y el 80 % de los casos (Barrera-López *et al.*, 2018) y es también una de las principales razones de incapacidad y dependencia progresiva en aquellos que la padecen. De acuerdo con cifras reportadas por la organización Alzheimer's Disease International, en el 2019 existían un poco más de 50 millones de personas con demencia a nivel mundial y se estima que para el año 2050 esta cifra aumente hasta 152 millones (Alzheimer's Disease International, 2019), debido al acelerado envejecimiento de la población mundial.

Siguiendo esta tendencia, en México también se espera un rápido crecimiento de la población de adultos mayores dentro de los próximos 20 años. Según estimaciones del Consejo Nacional de Población (CONAPO), del periodo comprendido del año 2000 al 2030, la población en el rango de 0 a 14 años de edad tendrá un decremento del 23 %, la población en el rango de 15 a 64 años sólo un incremento del 47 %, mientras que la población mayor a 64 años presentará un alarmante crecimiento del 300 % (Fernández *et al.*, 2014). Por lo que para el año 2030, el país tendrá más adultos mayores que jóvenes menores a 15 años, y para 2050, un 30 % de la población tendrá más de 60 años (Lizama, 2018).

El análisis de los datos anteriores resulta de interés debido a que dichas cifras revelan la importancia de mejorar los procedimientos, métodos y técnicas para el manejo de la salud de personas de la tercera edad. Así como la necesidad de desarrollar herramientas que puedan apoyar en tareas de diagnóstico, tratamiento y cuidado para los diversos problemas de salud que están presentes en este grupo poblacional; del cual casi una tercera parte serán adultos mayores dependientes, y una de las mayores causas de dependencia en la tercera edad es precisamente la demencia (Gutierrez, 2014).

Aunque actualmente la demencia no es curable, se busca que las personas con demencia (PcDs) puedan preservar una buena calidad de vida el mayor tiempo posible. En este sentido, existen tanto tratamientos farmacológicos como no farmacológicos, enfocados retrasar el deterioro cognitivo, así como a mitigar los síntomas psicológicos y de comportamiento en las PcDs. Inicialmente, es recomendable que los tratamientos partan de una intervención no farmacológica (InF), debido a que estos tratamientos promueven el uso de diferentes métodos y técnicas para proveer estabilidad física y emocional a los pacientes, sin los efectos secundarios que conllevan los tratamientos farmacológicos (Navarro *et al.*, 2016). A través de la implementación de entornos apropiados, tareas estimulantes y terapias aplicadas acorde a las necesidades específicas de cada paciente, el efecto de este tipo de intervenciones puede ser maximizado (Gitlin *et al.*, 2010).

Dentro de las InFs se encuentra la terapia de reminiscencia (TR), la cual se recomienda implementar en PcDs; particularmente a personas con demencia tipo Alzheimer (González-Arévalo, 2015). La TR se lleva a cabo mediante conversaciones entre la PcD y su cuidador, estas conversaciones son sobre memorias y experiencias agradables del pasado de la PcD, donde la PcD puede ser guiada cronológicamente a través de sus experiencias de vida o recuerdos particulares que evoquen emociones favorables (Woods *et al.*, 2018). En la TR se busca estimular la memoria autobiográfica usando información sobre su historia de vida, preferencias y hábitos, ya que se ha observado que este tipo de información se encuentra relativamente bien preservada y fácil de recordar por PcDs en una etapa inicial, ya que durante esta etapa se ve afectada la memoria a corto plazo principalmente (Alzheimer's Association, 2020).

En distintos trabajos se han analizado los beneficios de someter a las PcDs a este tipo de terapias y se observó que llevar a cabo de forma constante la TR en PcDs contribuye a mejorar sus estados psicológicos inestables, emocionales y

conductuales, aumentando su capacidad de socializar nuevamente, su autoestima y confianza (Olazarán *et al.*, 2010; Gonzalez *et al.*, 2015; González-Arévalo, 2015).

1.1. Problema de investigación

Ante el panorama mencionado previamente, se tiene un problema médico-social debido a un acelerado envejecimiento de la población y por lo tanto un crecimiento el padecimientos relacionados a este cambio demográfico. Debido a esto, surge la necesidad de desarrollar herramientas tecnológicas enfocadas a la asistencia en los cuidados a largo plazo de este grupo poblacional, en el que se incluyen a PcDs.

Se ha mencionado que como parte de los cuidados de PcDs es aconsejable llevar a cabo tratamientos no farmacológicos como la TR. Sin embargo, en la mayoría de los casos se vuelve casi imposible visitar a un especialista o terapeuta cada día, incluso para los cuidadores primarios el tiempo y esfuerzo gastado en brindar este tipo de terapia de forma periódica resulta bastante limitado debido a las muchas otras tareas en que tienen que apoyar.

En este sentido, se ha considerado relevante establecer como trabajo de investigación la creación de un modelo de diálogo que permita generar conversaciones personalizadas de acuerdo con la información de vida de la PcD. Este modelo de diálogo tiene el objetivo de apoyar a los cuidadores de PcDs a brindar TR de forma constante a pacientes con Alzheimer en etapa inicial.

1.2. Objetivos de la investigación

1.2.1. Objetivo general

Crear un modelo de diálogo para su integración en un prototipo de sistema conversacional, que permita una interacción humano-computadora mediante lenguaje natural enfocado en asistir a brindar terapia de reminiscencia a personas con Alzheimer en etapa inicial.

1.2.2. Objetivos específicos

- Implementar un mecanismo para representar la información de pacientes con Alzheimer.
- Crear un modelo de diálogo centrado en el usuario para generar conversaciones personalizadas.
- Integrar el modelo creado dentro de un prototipo de sistema conversacional.
- Establecer métricas que permitan evaluar la funcionalidad del modelo creado.

1.3. Preguntas de investigación

1. ¿De qué forma podría un sistema conversacional apoyar en brindar terapia de reminiscencia a una persona con Alzheimer en etapa inicial?
2. ¿Qué características requiere un modelo de diálogo para ser integrado a un sistema conversacional enfocado a brindar terapia de reminiscencia a personas con Alzheimer?
3. ¿Cómo puede ser evaluada la efectividad de un sistema conversacional que brinde terapia de reminiscencia a personas con Alzheimer?

1.4. Organización de la tesis

El documento está organizado en 6 capítulos, los cuales se describen a continuación. En el capítulo 1, se describe el problema a resolver, así como los objetivos planteados. El capítulo 2 introduce conceptos que son relevantes dentro del marco teórico para comprender la investigación realizada. Así, se describe inicialmente en que consiste la terapia de reminiscencia y posteriormente se definen algunos técnicas del área de procesamiento de lenguaje natural (PLN) implementados en esta investigación y finalmente se describe el concepto de sistema de diálogo.

En el capítulo 3 se realiza una revisión del estado del arte en el desarrollo de sistemas de diálogo enfocados a la salud y específicamente en el dominio de la demencia. Posteriormente en el capítulo 4 se expone el trabajo desarrollado a lo

largo de la presente investigación, siendo en la parte final del capítulo en donde se describe el modelo de diálogo propuesto y su integración al prototipo de sistema conversacional desarrollado.

Dentro del capítulo 5 se muestran los experimentos realizados para evaluar por una parte el desempeño de distintas herramientas de reconocimiento automático del habla y por otro lado la efectividad del prototipo de sistema conversacional al brindar la terapia.

En el capítulo 6 se discuten los aportes realizados durante el desarrollo de esta investigación, así como las principales conclusiones a las que se han llegado. Finalmente, se describen las oportunidades de trabajo futuro que se pueden continuar a partir de los resultados alcanzados.

Capítulo 2

Marco teórico

En este capítulo se exponen los conceptos y términos empleados a lo largo de este trabajo. Inicialmente, se detalla en que consiste la terapia de reminiscencia. A continuación se describen las principales características de un sistema de diálogo, así como su arquitectura y los principales elementos que la conforman. Finalmente, se describen las características del lenguaje AIML, el cual se utilizó para la generación de las respuestas del modelo de diálogo.

2.1. Terapia de reminiscencia

La terapia de reminiscencia es una de las intervenciones psicosociales más eficaces para llevarla a cabo en PcDs (Irazoki *et al.*, 2017). Es una forma de mantener el pasado del individuo y perpetuar su identidad a través de estimulación, comunicación y entretenimiento. La reminiscencia también se define como la generación de memorias, vivencias y eventos de nuestro pasado, usualmente el pasado distante (Dean~Webster, 2003). Por lo tanto, la TR involucra la discusión de actividades del pasado, eventos y experiencias con otra persona o grupo de personas.

Dentro del tratamiento de la demencia, la TR se considera una InF que busca promover el bienestar y mejorar la calidad de vida de las personas que padecen algún tipo de demencia como el Alzheimer. La TR ha mostrado una alta efectividad especialmente en mejorar el estado de ánimo, incrementando la sociabilidad y mitigando los síntomas de depresión de la PcD (González-Arévalo, 2015; Woods *et al.*, 2018).

La TR debe brindarse de acuerdo a las características de los participantes y su habilidad cognitiva. Los materiales o información usados en la TR deben de ser lo más relacionados posible a los pacientes; si la TR se brinda en grupo, este debe de ser pequeño y los participantes deberán tener un desempeño cognitivo semejante. Además, los familiares de los pacientes pueden ayudar a preparar las sesiones o participar en ellas brindando información personal del paciente. En caso de que el paciente exprese algunos recuerdos erróneos es recomendable escucharlo sin corregirlo, ya que el objetivo principal de la TR debería ser privilegiar la validación emocional sobre la corrección hacia los hechos reales (Gonzalez *et al.*, 2015).

2.2. Sistema de diálogo

Un sistema de diálogo (SD o DS, *Dialogue System*) es también conocido comúnmente como un sistema de interacción, entendiendo el diálogo como el proceso de comunicación entre varias partes que involucra el uso del lenguaje como mecanismo de interacción (Alonso-Martín *et al.*, 2013).

De forma similar a como se lleva a cabo el diálogo entre las personas, los sistemas de diálogo tratan de estructurar el diálogo en turnos. Cada turno está conformado por una o más emisiones de un participante en el diálogo. De tal forma que este intercambio de emisiones mediante turnos conforma el diálogo.

Por lo tanto, un sistema de diálogo hablado (SDH o SDS, *Spoken Dialogue System*) desde un enfoque computacional suele ser un sistema que incorpora un número diverso de módulos para gestionar de forma exitosa la interacción humano-computadora a través de lenguaje natural. Aunque existen diversas propuestas para el diseño, desarrollo e implementación de SDSs dependiendo del dominio de aplicación; un gran número de ellas se desarrolla a partir de una arquitectura clásica (Llisterri *et al.*, 2003), como la que se observa en la figura 2.1.

De acuerdo a esta arquitectura un SDS se compone esencialmente de cinco módulos principales: 1) Módulo de Reconocimiento Automático del Habla, 2) Módulo de Comprensión del Lenguaje Natural, 3) Módulo de Gestión del Diálogo (Gestor de diálogo), 4) Módulo de Generación de Lenguaje Natural y 5) Módulo de Conversión de Texto a Habla. Estos módulos llevan a cabo diversas tareas de PLN, de las cuales se brinda una descripción en las subsecciones siguientes.

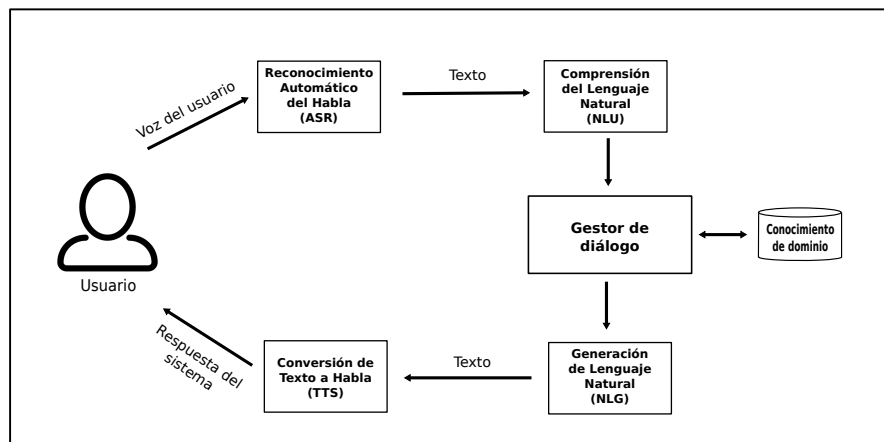


FIGURA 2.1: Arquitectura clásica de un sistema de diálogo

De forma general, el funcionamiento de un SDS se puede definir de la siguiente manera: inicialmente, el SDS recibe información por parte del usuario (voz del usuario). Posteriormente, el SDS realiza el reconocimiento automático del habla (ASR, *Automatic Speech Recognition*), convirtiendo las ondas de sonido que fueron recibidas (voz del usuario) a una secuencia de palabras. Como siguiente paso, el SDS analiza el texto transcrito y obtiene una interpretación de la información contenida en el mensaje, lo cual es realizado dentro del módulo de comprensión del lenguaje natural (NLU, *Natural Language Understanding*).

Entonces, mediante el gestor de diálogo (GD o DM, *Dialogue Manager*) el sistema puede reaccionar de manera adecuada: por ejemplo, si el usuario realizó una pregunta, el SDS puede tratar de encontrar una respuesta; si emitió una orden, entonces tratará de realizar la tarea. Para responder al usuario, generalmente el DM puede establecer comunicación con una base de datos local, buscar información en Internet o integrar un modelo de diálogo (dependiendo de las características y finalidad del SDS). Toda esta información es importante para que el DM pueda tomar una decisión enfocada a generar una respuesta satisfactoria al usuario.

Cuando el SDS está listo para brindar una respuesta de acuerdo a la decisión tomada por el DM, esta respuesta tiene que ser comprendida por el usuario; es decir, representarse de forma adecuada en lenguaje natural, tarea de la que se encarga el módulo de generación de lenguaje natural (NLG, *Natural Language Generation*) creando una representación textual de la respuesta. Por último se realiza la conversión de texto a habla de la respuesta empleando el módulo de conversión de texto a habla (TTS, *Text-to-Speech*).

A continuación se describen las tareas de PLN involucradas en cada uno de los módulos descritos.

2.2.1. Reconocimiento automático del habla

Se define como reconocimiento automático del habla al proceso de capturar y convertir a una representación textual una señal acústica recogida por un micrófono teniendo como fuente la voz humana (Alonso *et al.*, 2013). A consideración de Liddy (1998) y de Llisterri *et al.* (2003), en cualquier sistema de ASR existen los siguientes niveles lingüísticos de abstracción: nivel fonético, morfológico, léxico, sintáctico, semántico, de discurso, y pragmático. Estos niveles abarcan la construcción de palabras como una conjunción de fonemas y el conjunto de reglas válidas que determinan la posición de las palabras dentro de una oración.

De igual forma, en Alonso-Martín *et al.* (2013) mencionan que los sistemas de ASR pueden ser clasificados en dos tipos: el primero es “*speaker-dependent system*”, en el cual el sistema es entrenado para un usuario en específico, por lo que sólo es capaz de reconocer eficazmente a dicho usuario; este tipo de sistemas usualmente suelen ser herramientas de dictado. El otro tipo lo denominan “*speaker-independent system*” el cual es capaz de transcribir las oraciones habladas emitidas por cualquier usuario sin la necesidad de que el sistema sea entrenado para cada usuario en particular. La mayor parte de la investigación en el desarrollo de sistemas ASR se centra en el segundo tipo.

El problema de ASR puede ser visto como un caso especial de reconocimiento de patrones, el cual tiene el objetivo de extraer, analizar, modelar y reconocer información sobre un problema de clasificación planteado (Becerra *et al.*, 2017). Los sistemas de ASR más recientes modelan la variabilidad de la señal del habla empleando modelos ocultos de Markov (HMM, *Hidden Markov Model*) y algoritmos de aprendizaje profundo (Miao *et al.*, 2015; Wu *et al.*, 2017)

2.2.2. Comprensión del lenguaje natural

La comprensión del lenguaje natural se clasifica como un subtópico del área de PLN que actualmente sigue siendo considerado un problema difícil de resolver en el campo de la AI (Yampolskiy, 2013). El término NLU puede ser utilizado para

referirse a un conjunto diverso de aplicaciones computacionales, que van desde tareas pequeñas relativamente simples como el entendimiento de comandos cortos emitidos hacia un robot, hasta tareas sumamente complejas como la comprensión del texto de un documento o el entendimiento de la emisión hablada de un usuario en un sistema de diálogo por habla. Por lo tanto, dentro de los objetivos del estudio de NLU se busca realizar la extracción de información desde texto o habla, para entonces poder crear una descripción de los conceptos representados en cualquiera que sea el medio utilizado (Weik, 2001).

Independientemente del enfoque utilizado al desarrollar sistemas que implementen en un cierto nivel NLU, la mayoría de estos sistemas comparten algunos componentes en común los cuales son: un lexicón¹ del lenguaje, un analizador sintáctico y reglas gramaticales para descomponer las oraciones en una representación interna que pueda ser analizada por el sistema. La construcción de un lexicón apropiado a un lenguaje o contexto requiere un esfuerzo significativo, ya que se necesitan emplear una gran cantidad de recursos humano-tiempo para construirlo (Miller, 1995). Por lo que actualmente, métodos basados en técnicas de aprendizaje profundo comienzan a representar una opción viable para la creación de modelos analizando una gran cantidad de datos (Devlin *et al.*, 2018; Burtsev *et al.*, 2018).

Es necesaria una teoría semántica que sirva como base para la comprensión del lenguaje, ya que las capacidades de interpretación o entendimiento del sistema dependerá de la teoría semántica empleada. Las teorías semánticas competentes del lenguaje tienen ventajas y desventajas específicas en cuanto a su idoneidad como base de la interpretación semántica automatizada por computadora (Lawler y Dry, 2008). Recientemente, propuestas de aplicaciones de NLU también intentan incorporar inferencia lógica dentro de su funcionamiento, para ello hacen un mapeo de la representación obtenida en un conjunto de cláusulas dentro de la lógica de predicados, para entonces usando deducción lógica obtener conclusiones a partir de las cláusulas definidas (Pan *et al.*, 2019).

2.2.3. Gestor de diálogo

El gestor de diálogo representa una parte esencial de un SDS, ya que se encarga de tareas primordiales como:

¹Se define al lexicón como el vocabulario de una persona, lenguaje o rama del conocimiento.

- Mantener el estado de la conversación.
- Gestionar el flujo de interacción entre el usuario y el SDS.
- Esperar por una emisión del usuario.
- Gestionar el historial del diálogo.
- Información contextual y adquisición de distintas fuentes de conocimiento si es necesario.

Algunas tareas típicas realizadas por el DM se relacionan a la toma de decisiones por parte del SDS, como pueden ser: realizar una tarea o acción específica (al recibir un comando por parte del usuario), brindar información al usuario (en respuesta a una pregunta), solicitar información al usuario (mediante la generación de una pregunta), realizar acciones de verificación o confirmación en caso de incertidumbre o de información incompleta para continuar con la interacción.

Una importante tarea que se resuelve entre el módulo NLU y el DM es la de poder identificar y clasificar de los actos de habla, los cuales expresan o determinan la función de una emisión. El análisis de los actos de habla involucra más que un análisis sintáctico y semántico, ya que se pueden basar en información adicional como el contexto de la conversación (Gunkel, 2016). Debido a que la finalidad de una emisión se relaciona con el objetivo del usuario, el análisis de los actos de habla juega un papel importante en la toma de decisiones del DM (O'Shea *et al.*, 2012).

Existen diferentes enfoques para el diseño del gestor de diálogo, los cuales se implementan de acuerdo al dominio de aplicación en el que se utilizará el sistema de diálogo. Los diferentes enfoques que han sido propuestos y pueden ser encontrados en la literatura para la administración del diálogo se pueden clasificar de acuerdo a su función en dos clases, las cuales son: la estrategia de control del diálogo y el modelo de control del diálogo.

2.2.3.1. Estrategias de control de diálogo

Las estrategias de control del diálogo consideran en qué medida el usuario y/o el sistema controlan la iniciativa en un diálogo (Voorhees, 2003; del Valle-Agudo *et al.*, 2006). Las posibles estrategias que pueden ser implementadas son:

- *Basados en la iniciativa por parte del SDS.* En este tipo de estrategia el SDS es quien guía el diálogo y realiza preguntas para obtener información del usuario, el cual solo puede interactuar con el sistema únicamente respondiendo a las preguntas realizadas.
- *Basados en la iniciativa por parte del usuario.* En este caso, la estrategia del diálogo es llevada a cabo por el usuario y puede realizar preguntas al SDS, el cual es capaz de generar una respuesta, o ejecutar alguna tarea establecida.
- *Basados en iniciativa mixta.* En este caso las estrategias anteriores son combinadas y tanto el SDS como el usuario pueden tomar la iniciativa. Es decir, el usuario puede realizar una pregunta en cualquier momento del diálogo, pero el SDS también puede asumir el control del diálogo para obtener información del usuario.

2.2.3.2. Modelos de control del diálogo

Las estrategias basadas en iniciativa por parte del SDS y de iniciativa mixta requieren que el SDS implemente un modelo de control con el objetivo de poder gestionar el flujo del diálogo. Los principales enfoques en la gestión y control del diálogo son: modelos basados en estados finitos (POMDP), modelos de estado de información (AIML), enfoque basado en marcos y enfoque basado en agentes (Mavridis, 2015; Mairittha *et al.*, 2019).

Dentro de este trabajo se ha establecido emplear un modelo de control del diálogo basado en estados de información, por lo que se implementó el lenguaje AIML para la generación de bases de conocimiento del modelo de diálogo.

2.2.4. Generación de lenguaje natural

La generación de lenguaje natural se define como una tarea de producir texto en lenguaje natural entendible para los humanos a partir de sistemas de representación textuales (Perera y Nand, 2017). Dentro del desarrollo de SDSs, la tarea de NLG puede ser vista como un problema de cómo decir algo una vez que el SDS ha determinado qué es lo que se quiere decir.

La tarea de NLG puede en este sentido ser vista como lo opuesto a NLU, ya que mientras que para NLU el sistema trata de desambiguar el texto de entrada para

producir un interpretación que pueda ser analizada por la máquina, en NLG el sistema necesita tomar decisiones sobre como representar un concepto con palabras en lenguaje natural. De tal forma que el proceso de NLG en un sistema de diálogo abarca un gran rango de tareas que van desde la estructuración del contenido, la elección de las expresiones a utilizar y hasta la elección de marcadores de entonación para la producción del habla (Lemon, 2011). Por lo tanto, los sistemas sofisticados que implementan NLG necesitan incluir etapas de planeación y unión de información para poder generar texto que luzca natural y no sea repetitivo.

La generación del mensaje textual debe de ser coherente con el actual estado del diálogo y su construcción requiere identificar: 1) la información que debe de ser incluida, 2) cómo debe de ser estructurada esta información y 3) la elaboración lingüística del mensaje; es decir, la formulación del mensaje, incluyendo la selección de elementos léxicos y estructuras sintácticas (Russo *et al.*, 2019).

2.2.5. Conversión de texto a habla

Dentro de la tarea de conversión de texto a habla se busca transformar (sintetizar) oraciones de texto en habla. En el caso de las propuestas basadas en diccionarios restringidos, palabras u oraciones pre-grabadas pueden ser concatenadas para producir una emisión de salida. Sin embargo, herramientas actuales permiten transformar texto arbitrario en habla. Pero hasta este momento la síntesis de texto a habla continua siendo una tarea compleja, ya que esta conversión es precedida por una etapa de análisis de texto que incluye la segmentación y normalización del texto, análisis morfológico, determinación de la entonación y el modelado de un efecto de habla continua con el objetivo final de producir un resultado que se asemeje lo más posible a una voz real (Xu y Lee, 2017).

2.3. Clasificación de los sistemas de diálogo

Si bien actualmente existe un gran interés en el desarrollo de sistemas de diálogo desde distintos enfoques. De forma general los sistemas de diálogo se suelen clasificar de acuerdo a su objetivo en tres categorías: i) Sistemas de diálogo enfocados a tarea, ii) Sistemas de diálogo de pregunta-respuesta y iii) Sistemas de diálogo conversacionales.

2.3.1. Sistemas de diálogo enfocados a tarea

Este tipo de SDs se enfocan en realizar una tarea bien definida a partir de las instrucciones o comandos recibidos por parte del usuario. Por lo regular el diálogo presenta una estructura bien definida a partir de la funcionalidad que se pretende dar al SD. Los objetivos o tareas a cumplir por el SD comúnmente se refieren a realizar alguna acción, recuperar información o consultar bases de datos. Por lo que existe un gran campo de aplicación de este tipo de SD, como lo son: hogares inteligentes, sistemas multimedia por voz, asistentes inteligentes, sistemas para reservación y venta de boletos, entre otros.

Un SD enfocado a la tarea suele tener un comportamiento de iniciativa mixta, ya que si bien el usuario es el quien inicia el diálogo estableciendo su objetivo, el sistema puede tener una participación mayor durante el diálogo ya que necesita obtener información suficiente para realizar la tarea deseada por el usuario. Sin embargo, lo que se busca en este tipo de sistemas es satisfacer el objetivo con el menor número de interacciones posibles.

Una estrategia ampliamente usada para recabar toda la información que el sistema necesita para lograr su objetivo es modelar la interacción como una secuencia de acciones que necesita realizar el sistema. A estas acciones se les denomina actos de habla o intención de la emisión. Cada uno de los actos de habla esta definido por un tipo y por los argumentos que puede recibir. Por lo que el uso de ciertos actos de habla dependerá del modelado del diálogo implementado. En este sentido Young *et al.* (2010) proponen un conjunto de actos de habla para la gestión del diálogo en sistemas de diálogo enfocados a tarea, basándose en las categorías de los actos de habla asertivos, directivos y expresivos descritos en la taxonomía propuesta por Searle (1969). Cada emisión del usuario es representada mediante un acto de habla y los actos de habla generados por el sistema son convertidos a una emisión en lenguaje natural.

2.3.2. Sistemas de diálogo de pregunta-respuesta

Este tipo de sistemas presenta similitud con los SDs enfocados a tarea, ya que si bien no se espera que el sistema realice una acción en específico, en este caso la tarea podría ser definida como buscar la respuesta correcta a una pregunta realizada por el usuario.

De forma general, un SD de pregunta-respuesta permite al usuario obtener o buscar información mediante lenguaje natural, brindando una respuesta concreta a la pregunta del usuario (Singh *et al.*, 2020). Los sistemas de pregunta-respuesta pueden ser clasificados a su vez en: sistemas de turno único, basados en contexto e interactivos. Siendo los primeros, los más comunes, en este caso, el SD es desarrollado para brindar una respuesta única al usuario sin ninguna interacción adicional. Este tipo de sistemas funciona bien para preguntas factuales (Voorhees, 2008). Sin embargo, presentan dificultades al manejar preguntas complejas, las cuales requieren varios turnos de interacción para poder inferir la respuesta.

En este caso, los sistemas basados en contexto presentan un mejor desempeño, ya que llevan un registro de las preguntas realizadas previamente, lo que permite resolver ambigüedades (por ejemplo, cuando se hacen uso de pronombres para sustituir el uso de entidades). Por lo que la respuesta del sistema no depende únicamente de la última pregunta realizada, sino también del contexto del diálogo dado por preguntas y respuestas previas. Debido a esto, a este tipo de sistemas también se les denomina SDs de pregunta-respuesta multi-turno (Choi *et al.*, 2018).

Por último, los sistemas pregunta-respuesta interactivos utilizan algunas características de los sistemas enfocados a tarea. Esto debido a que si una pregunta no puede ser respondida entonces el sistema realiza la interacción con el usuario para sugerir ampliar los términos de búsqueda, o por el contrario si se obtienen varios resultados restringir o limitar estos términos para reducir la lista de resultados (Rieser y Lemon, 2009).

Actualmente existen sistemas que integran características tanto de los SDs enfocados a tarea como de los de pregunta respuesta. Estos sistemas denominados asistentes inteligentes (Siri ², Alexa ³, Google Assistant ⁴) combinan las capacidades de realizar una tarea (reproducir música, controlar dispositivos, realizar reservaciones, etc.) con las de buscar información específica que el usuario requiere (clima, noticias, direcciones, horarios, servicios, etc.). Sin embargo, no cuentan con la habilidad de mantener conversaciones sobre un dominio abierto con los usuarios.

²<https://www.apple.com/siri/>

³<https://alexa.amazon.com/>

⁴<https://assistant.google.com/>

2.3.3. Sistemas de diálogo conversacionales

Los SDs conversacionales son diseñados para llevar a cabo conversaciones no estructuradas y de dominio abierto o al menos más amplio que el de los sistemas enfocados a tarea o de pregunta-respuesta.

Si bien la intención de este tipo de sistemas es emular el comportamiento humano, no necesariamente en todos los casos se busca que el usuario tenga la percepción de que está conversando con otra persona. Además, de forma similar a los SDs enfocados a tarea, los SDs conversacionales constan de un comportamiento de iniciativa mixta. Esto debido a que su objetivo se centra en mantener una conversación el mayor tiempo posible con el usuario en lugar de tener que realizar o cumplir con una tarea en específico (Zhou *et al.*, 2020).

De forma general existen dos tipos de enfoques para el desarrollo de SDs conversacionales: sistemas basados en reglas y sistemas basados en corpus. Por un lado los sistemas basados en reglas están basados en un conjunto de reglas que determina su comportamiento. Para hacerlo se emplean reconocimiento de patrones y reglas de transformación, por lo que se toma la emisión del usuario y se aplican una serie de transformaciones sobre ella para generar o buscar una respuesta. Por otro lado, los sistemas basados en corpus emplean grandes cantidades de datos para modelar una estrategia de diálogo que pueda ser utilizada dentro del sistema conversacional. Esto permite entrenar agentes conversacionales usando técnicas de aprendizaje profundo (Jadhav y Thorat, 2020). La principal ventaja de los sistemas basados en reglas es que brindan un buen desempeño en dominios específicos, pero no para un dominio abierto. Mientras que los sistemas basados en corpus ofrecen la habilidad de funcionar sobre un dominio más amplio siempre y cuando exista una gran cantidad de datos para poder ser entrenados.

Algunos ejemplos de implementaciones de estos sistemas son en el desarrollo de humanos virtuales, chatbots sociales o agentes conversacionales. Los principales dominios para los cuales se han desarrollado este tipo de sistemas son el de la salud y el de la educación. En el dominio de la salud enfocados al diagnóstico (Tanaka *et al.*, 2017), apoyo en tratamientos (Fitzpatrick *et al.*, 2017) y en el monitoreo y rehabilitación de pacientes (Kamita *et al.*, 2019; Inkster *et al.*, 2018). Mientras que en el dominio de la educación se centra en el desarrollo de tutores inteligentes (Sosnowski y Yordanova, 2020; Lippert *et al.*, 2020).

2.4. Métricas de evaluación de sistemas de diálogo conversacionales

Si bien es necesario evaluar el desempeño de los SDs, esta tarea resulta un poco más compleja al evaluar sistemas de diálogo conversacionales, debido a que la definición de lo que se considera un diálogo adecuado o apropiado no siempre es clara y a menudo depende del dominio de aplicación del SD. Por lo que, asumiendo que la calidad del diálogo depende de la habilidad del sistema de brindar respuestas adecuadas, no siempre es claro como evaluar la idoneidad de las respuestas para cada sistema en particular (Deriu *et al.*, 2021).

Esto representa un reto durante la etapa de evaluación de un SD conversacional. Ya que a diferencia de los SDs enfocados a tarea y pregunta-respuesta donde principalmente se evalúa el desempeño del sistema para realiza una tarea o responde una pregunta de forma correcta; en los SDs conversacionales se busca evaluar diferentes características como: la similitud al comportamiento humano, variabilidad e idoneidad de las respuestas (Gandhe y Traum, 2016), diversidad de tópicos de conversación, coherencia en el diálogo, entre otras.

Y aunque el método de evaluación más común y con mayor precisión sigue siendo mediante evaluadores humanos, en muchos casos esto resulta bastante costoso en cuanto a recursos económicos, de tiempo y esfuerzo. Por lo que con el objetivo de contar con métodos que permitan realizar una evaluación automática y estandarizada de este tipo de SDs se han propuesto distintas métricas de evaluación para las diferentes características de un sistema conversacional.

Inicialmente se han tratado de adoptar métricas de evaluación automática desarrolladas para otras tareas de PLN. Por ejemplo, BLEU (Papineni *et al.*, 2002) y METEOR (Banerjee y Lavie, 2005), las cuales ya son un estándar para la evaluación de modelos de traducción automática, así como ROUGE (Lin, 2004) usada en evaluación de resúmenes automáticos. Sin embargo, estas métricas asumen que respuestas válidas o adecuadas tienen una significativa superposición de palabras con las respuestas muestra o esperadas. Pero esto no se puede considerar como un caso general en los SDs conversacionales, donde existe una diversidad de respuestas que se pueden considerar adecuadas para un contexto dado (Liu *et al.*, 2016). Una propuesta desarrollada por Galley *et al.* (2015) denominada delta BLEU propone

incluir el juicio humano dentro de la métrica de BLEU en donde evaluadores humanos califican las respuestas de referencia estableciendo su relevancia de acuerdo al contexto. Sin embargo, analizando la correlación entre los resultados de estas métricas y el juicio humano para evaluar la calidad de las respuestas, esta suele ser muy baja o nula (Liu *et al.*, 2016), por lo que no se aconseja el uso de estas métricas para la evaluación de SDs conversacionales.

Por otro lado, Gandhe y Traum (2016) abarcan el problema de tener múltiples emisiones (utterances) como respuestas candidatas y proponen una métrica que toma esto en cuenta denominada *Voted appropriateness* (VA) (Idoneidad votada). VA es una mejora de la métrica de *Weak agreement* (WA) (Acuerdo débil) (DeVault *et al.*, 2011) la cual se basa en la observación de que los evaluadores humanos por lo regular solo coinciden el 50 % de los casos al seleccionar una respuesta del sistema como adecuada dado un contexto. Los autores atribuyen esto al hecho de que múltiples respuestas pueden ser consideradas como adecuadas o idóneas. Por lo que la métrica de WA considera una respuesta como apropiada si al menos un anotador humano selecciona la respuesta como tal.

A diferencia de WA que considera equitativamente cada respuesta seleccionada como correcta, la métrica de VA toma en cuenta el número de evaluadores que seleccionaron una respuesta como correcta para un determinado contexto. Esta métrica está definida por la ecuación 2.1, donde $V(u_t, context_t)$ representa el número de evaluadores que seleccionaron la emisión u_t como una respuesta apropiada al contexto de diálogo $context_t$. Modelando VA o R_{voted} como una regresión lineal basada en los valores de V .

$$R_{voted}(u_t, context_t) = \alpha_0 + \alpha_1 \cdot V(u_t, context_t) \quad (2.1)$$

Esto permite obtener una mejor correlación entre el juicio humano y las respuestas del sistema seleccionadas como apropiadas. Sin embargo, ambas métricas recaen en la evaluación de anotadores humanos, por lo que si bien tienen un buen desempeño al evaluar sistemas basados en selección de respuesta, se vuelen difíciles de implementar para evaluar sistemas conversacionales basados en corpus.

2.5. AIML

El propósito del lenguaje AIML (*Artificial Intelligence Markup Language*) es proveer un lenguaje enfocado a la tarea de modelar diálogos. Además, es un lenguaje basado en XML (*eXtensible Markup Language*) y está basado en etiquetas. Las etiquetas son identificadores responsables de crear fragmentos de código e insertar comandos dentro del agente conversacional. AIML define una clase de objeto de datos denominada objetos AIML, que se implementa en el modelado de patrones de conversación. Técnicamente hablando, los objetos AIML son etiquetas, y cada etiqueta corresponde a un comando dentro del lenguaje. La forma general de un comando o etiqueta AIML tiene la siguiente estructura:

<comando> lista de parámetros </comando>

Un comando en AIML consiste de una etiqueta de apertura <comando>, una etiqueta de cierre </comando> y el texto que contiene la lista de parámetros del comando. AIML es un lenguaje interpretado, por lo tanto cada sentencia es leída, interpretada y ejecutada por un software interprete.

AIML se basa en unidades básicas de diálogo, conformadas por patrones de entrada (del usuario) y respuestas (del modelo de diálogo), estas unidades básicas son conocidas como categorías y el conjunto de todas las categorías establecidas en el modelo forman la base de conocimiento del sistema o agente conversacional. Dentro de los objetos principales en AIML se encuentran la categoría (*category*), los patrones (*pattern*) y las plantillas (*template*). La etiqueta de *category* define una unidad de conocimiento, la etiqueta *pattern* define una posible entrada del usuario y la etiqueta *template* establece una respuesta o conjunto de respuestas para un determinado patrón de entrada. Con AIML es posible crear interfaces de comunicación humano-computadora en lenguaje natural manteniendo una respuesta rápida y la implementación altamente escalable. Además de que se posee de una jerarquía de etiquetas bien definida dentro de los objetos AIML creados como se muestra en la figura 2.2.

Otra característica de AIML es que permite el desarrollo de reglas y patrones de acuerdo al dominio de aplicación del sistema de diálogo. Esto permite ajustar dichos patrones para lograr una mayor precisión en la detección de estos patrones en las oraciones de entrada. Recientemente expertos han propuesto que los SDS

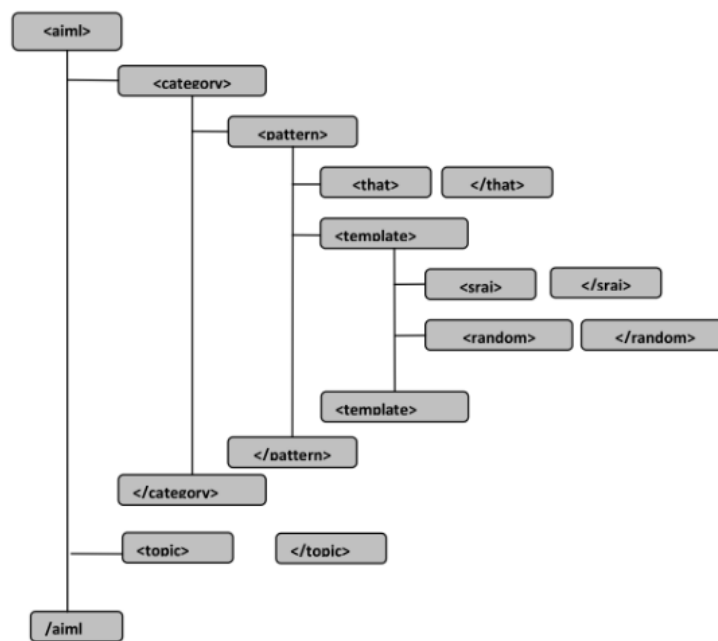


FIGURA 2.2: Jerarquía de etiquetas en objetos AIML

deberían de cumplir con objetivos de más alto nivel, tales como proveer al sistema de razonamiento avanzado (mediante el uso de ontologías, bases de conocimiento y sistemas de aprendizaje automático), facultad de adaptación, pro-actividad y multilingüismo. Sin embargo, debido a la ambigüedad y complejidad propia del lenguaje; hasta este momento no existe una propuesta que haya desarrollado un SDS de propósito general que pueda cumplir con todos estos requerimientos.

2.6. Conclusión del capítulo

Los conceptos y términos presentados en este capítulo permitirán ofrecer una mejor comprensión del trabajo desarrollado en esta tesis. Se han descrito principalmente aquellos referentes a la terapia de reminiscencia y los componentes que integran un sistema de diálogo. Además, la clasificación de los sistemas de diálogo de acuerdo a su objetivo, así como los principales métodos para evaluar los sistemas conversacionales.

Capítulo 3

Estado del arte

En este capítulo se describen las propuestas más relevantes en el desarrollo de sistemas de diálogo o interactivos enfocados a la salud, específicamente en el dominio de la demencia. Primero, se describen trabajos iniciales en el modelado del diálogo. Posteriormente se realiza un análisis de los trabajos enfocados en el desarrollo de sistemas de diálogo en el dominio de la demencia orientados al diagnóstico, apoyo cognitivo y a brindar terapia. Por último se discuten las limitantes y áreas de oportunidad resultantes de la revisión del estado del arte.

3.1. Trabajos iniciales en el modelado del diálogo

En estos días existe un gran interés por parte de la comunidad en investigar sobre la creación de sistemas de diálogo que permitan una interacción más natural entre humano-máquina través de lenguaje natural. Sin embargo, este interés no es tan nuevo, ya que décadas atrás se comenzaron a desarrollar propuestas para tratar de caracterizar el diálogo (Austin, 1968) y el estudio de la comunicación verbal entre humanos (Searle, 1969, 1975; Bach y Harnish, 1979), donde definen que los *actos comunicativos* o *actos del habla* son las unidades básicas del diálogo, los cuales pueden ser clasificados en afirmaciones, preguntas, directrices, respuestas, declaraciones, entre otras (Marineau *et al.*, 2000).

Un trabajo inicial en la creación de modelos de diálogo que se basan en la identificación y clasificación de los actos de habla, fue el presentado por Grosz y Sidner (1986). En el describen los elementos básicos de una teoría computacional para

tratamiento de la estructura del discurso. Para esto, se analizan las dos características del discurso: intención y atención. También afirman que la estructura del discurso consta de tres componentes: la estructura de secuencias de unidades del habla conocidas como *utterance* (término en inglés), la estructura de las intenciones y el estado de atención. Aunque el enfoque propuesto se centra en el análisis del discurso, el modelo que ellos proponen se puede aplicar a diálogos orientados a la tarea, donde uno de los interlocutores pretende que los demás participantes del diálogo realicen una acción.

Por otro lado, en Zubizarreta-Aizpuru (1993) proponen en el proceso de caracterización del diálogo, el empleo de funciones que usan los constituyentes del diálogo (interlocutores), denominadas funciones ilocutivas¹, las cuales cambian la relación entre los participantes del diálogo (Yáñez, 2013). Estas funciones se asocian con cada acto de habla y su uso puede requerir que se cumplan condiciones (precondiciones o poscondiciones) al interior del diálogo.

Adicionalmente al modelado del diálogo, también la gestión del diálogo ha sido ampliamente estudiada en sistemas de inteligencia artificial desde hace varias décadas atrás, por lo que analizando tales sistemas se pueden considerar de forma general dos aproximaciones diferentes: los gestores de diálogo basados en patrones, los cuales no tienen en cuenta el significado semántico de las oraciones, como por ejemplo *ELIZA* (Weizenbaum, 1966), *ALICE* (Wallace, 2001); y por otro lado los gestores que trabajan usando algún tipo de razonamiento semántico o teoría del diálogo natural, como por ejemplo el desarrollado para *TRINDI* (Larsson *et al.*, 1999).

Sin embargo, diferentes propuestas (Grosz y Sidner, 1986; Reilly *et al.*, 1988; Zubizarreta-Aizpuru, 1993) se basan únicamente en la caracterización del diálogo, pero no proponen la forma de etiquetar en un corpus hablado los actos de habla, ni la representación del modelo de diálogo según estas etiquetas, ni el mecanismo para gestionar un diálogo cualquiera.

En el etiquetado de los actos de habla, en Martínez-Hinarejos *et al.* (2002) proponen un conjunto de etiquetas para los actos de habla de tres niveles. El primer nivel de etiquetado es independiente del dominio del diálogo y describe su comportamiento (inicio-diálogo, cierre-diálogo, pregunta, respuesta, confirmación, etc.). El

¹El acto ilocutivo se refiere a la intención de decir algo. Searle (1969) lo desarrolla con el término de *fuerza ilocutiva*. La “fuerza” se refiere a la función comunicativa, si es mandar, preguntar, afirmar, etc.

segundo nivel se relaciona con la representación semántica de una frase hablada que realiza un interlocutor en un turno del diálogo. Finalmente, el tercer nivel representa la información proporcionada en un turno del diálogo que hace referencia a la ontología del dominio. A partir de esta clasificación, es posible construir un modelo estocástico según los actos de habla reconocidos, con el fin de representar el modelo de diálogo. Este modelo estocástico, tiene gran utilidad en la elaboración del gestor de diálogo.

Existen otras propuestas diferentes para el reconocimiento y etiquetado de los actos de habla, por ejemplo, usando redes estocásticas o máquinas de estados (Stent, 2000; Clark y Popescu-Belis, 2004; Litman *et al.*, 2006; Ji y Bilmes, 2006; Fernández *et al.*, 2007).

3.2. Sistemas de diálogo en el dominio de la demencia

Actualmente existe un gran interés en el desarrollo de sistemas interactivos o sistemas de diálogo que puedan ser empleados como herramientas de apoyo a médicos, cuidadores y pacientes dentro del dominio de la demencia. Estos abarcan diversos escenarios de aplicación como: la detección y diagnóstico de la demencia de forma prematura, el desarrollo de asistentes cognitivos inteligentes para el apoyo en la realización de tareas cotidianas y aquellos enfocados en apoyar a brindar terapia psicológica y emocional para ayudar a preservar la calidad de vida de los PcDs. La tabla 3.1 resume algunos de los trabajos más significativos dentro de cada uno de estos enfoques, los cuales se describen en a continuación.

3.2.1. Sistemas enfocados al diagnóstico

Un gran número de trabajos encontrados dentro de la literatura se enfocan en tratar de detectar y diagnosticar la demencia en una etapa temprana, lo que ofrecería a los PcDs y sus familiares establecer un tratamiento adecuado y mantener la calidad de vida del paciente por un tiempo mayor que cuando se realiza un diagnóstico tardío.

Primer autor	Enfoque del estudio	Tipo de sistema	Hallazgos principales
Tanka et al.	Diagnóstico	Avatar virtual	Alcanza una precisión del 93 % en la detección de demencia.
Chinaei et al.	Detección de confusión	Sistema de diálogo	Detecta confusión en el habla alcanzando 82 % de precisión
Yasuda et al.	Reminiscencia remota y asistencia en tareas	Videollamada	Concluyen que la estabilidad psicológica persiste por 3 horas después de la intervención
Rudzicz et al.	Asistencia en tareas de la vida diaria	Robot móvil	Encontraron que altos niveles de confusión en PcDs motivaban que ignoraran al robot (hasta 40 %)
Wolters et al.	Asistente cognitivo	Sistema de dialogo hablado	Resaltan que el tipo de voz y el estilo de de interacción deben ser establecidos basados en las preferencias del paciente y no del cuidador
Casey et al.	Apoyo cognitivo	Robot de compañía	Encontraron que aspectos relevantes para mejorar la aceptación por parte de los pacientes son: hablar lento y con tono de voz alto, personalizar la conversación a los intereses y preferencias del paciente
Navarro et al.	Terapia ocupacional	Aplicación de computadora	Peronalizando la intervención ayudo a facilitar la adopción de este tipo de sistemas
Cruz-Sandoval et al.	Terapia de estimulación cognitiva	Robot social	Observaron un decremento significativo en comportamientos erráticos durante y después de las sesiones

TABLA 3.1: Estudios centrados en el desarrollo de sistemas interactivos en el dominio de la demencia

Para abordar este problema, en el trabajo desarrollado por Tanaka *et al.* (2017) crean un avatar con funcionalidades de diálogo hablado limitadas que realiza preguntas al usuario para llevar a cabo una evaluación del estado mental. Las preguntas hechas a los PcDs están basadas en la escala de memoria de Wechsler, el Mini-Examen del Estado Mental (MMSE, *Mini-Mental State Examination*)² (Cockrell y Folstein, 2002) y distintos cuestionarios neuropsicológicos validados. La conversación entre el avatar y el PcD es grabada y posteriormente realizan la extracción de diversas características audiovisuales, las cuales son empleadas dentro de dos algoritmos de aprendizaje automático (SVM y LR) con el fin de poder predecir prematuramente signos de demencia analizando diversas características del habla y comportamiento de los PcDs, manifestando una precisión de 0.93 en la detección.

En la propuesta desarrollada por Chinaei *et al.* (2017) analizan distintas características lingüísticas que son indicadores verbales de confusión en pacientes con

²El MMSE es un examen ampliamente usado dentro del dominio médico que evalúa la función cognitiva de adultos mayores. Los aspectos que se consideran dentro del examen son: orientación, atención, memoria, habilidades de lenguaje y espaciales.

Alzheimer, entre estas características se encuentran la riqueza del vocabulario, el análisis de la estructura del árbol sintáctico y las señales acústicas. Haciendo uso de diversos algoritmos de aprendizaje automático logran alcanzar una precisión de hasta 82 % en la detección de la confusión. Ellos proponen implementar estrategias de diálogo para evitar la confusión en la comunicación utilizando POMDP (*Partially Observable Markov Decision Process*).

3.2.2. Sistemas enfocados al apoyo cognitivo

El objetivo de los asistentes inteligentes o asistentes cognitivos es apoyar a PcDs a realizar actividades de la vida diaria detectando cuando ocurre un problema y ofreciendo asistencia personalizada y sensible del contexto (Wolters *et al.*, 2016). La implementación de un SDS como un parte de la arquitectura de los asistentes brinda la posibilidad de interactuar con los PcDs a través de la voz, permitiéndoles enfocarse en la tarea a realizar y no en la manipulación del asistente. La mayoría de propuestas se centra en el desarrollo de robots sociales o asistentes virtuales como las que se describen a continuación.

En Rudzicz *et al.* (2015) emplean un robot asistencial (ED), diseñado para ayudar en las tareas cotidianas a través de monitoreo visual y recordatorios por habla. Ellos analizan las interacciones basadas en habla entre ED y cada uno de los 10 pacientes con Alzheimer involucrados en este estudio. Como resultado de su investigación, pudieron observar que en las partes del diálogo donde el paciente mostró confusión, simplemente ignoró las indicaciones del robot, lo cual ocurrió el 40 % del total de las interacciones con el robot. Sin embargo, los participantes manifestaron una aceptación general del sistema sugiriendo varias mejoras para su implementación, entre las que destacan: 1) tomar en cuenta el estado emocional y mental del paciente y 2) mejorar las funciones de habla del robot, ya que para ellos resultó mucho más fácil comunicarse a través de la voz que seleccionar opciones en la pantalla de ayuda.

Por otro lado, en el trabajo desarrollado por Wolters *et al.* (2016) realizaron un análisis sobre 3 diferentes grupos de estudio (PcDs, cuidadores y personas mayores sin demencia) con el objetivo de recabar los requerimientos necesarios para el diseño de un asistente cognitivo de ayuda para PcDs. Ellos emplean un asistente simulado para interactuar con los usuarios y ayudarlos a utilizar una aplicación de calendario. El asistente brinda la sensación de contar un sistema de diálogo

intencional, es decir, el sistema guía al usuario a lo largo del diálogo. Algunos resultados relevantes que obtienen son: que tanto las PcDs como sus cuidadores reaccionaron positivamente a que se pueda establecer la comunicación por voz, pero las personas mayores sin demencia manifestaron sentirse muy guiados o limitados en la interacción. También se observó que la mayoría de PcDs percibían al asistente como compañía más que como asistente para una tarea específica, por lo que realizaban una interacción más social, lo que sobrepasaba las capacidades del componente de NLU al tener un vocabulario bastante limitado.

3.2.3. Sistemas enfocados a terapia

Los trabajos agrupados en esta categoría se enfocan en apoyar a los cuidadores (enfermeros o familiares) a contar con distintas herramientas o métodos que permitan retardar el deterioro cognitivo y mitigar los síntomas psicológicos, emocionales y de comportamiento que presentan las personas con algún tipo de demencia. Algunas propuestas relacionadas al trabajo desarrollado en esta investigación se describen a continuación.

Un trabajo inicial en enfocarse en brindar terapia de reminiscencia empleando un sistema de video llamadas para mejorar la estabilidad emocional de PcDs es el desarrollado por Yasuda *et al.* (2013), en donde además brindan recordatorios y asistencia para realizar sus tareas cotidianas. Los autores observaron que la estabilidad psicológica de un paciente persistía hasta por un lapso de tres horas posteriores a una intervención. Además, después de que fueron agregados a la interacción algunos videos que mostraban como realizar las tareas cotidianas, ellos encontraron que el nivel de completar las tareas en comparación con pacientes que no usaron el sistema fue un 82 % mayor. Sin embargo, la interacción siempre se realiza entre paciente y un cuidador humano, lo que limita la constancia en la periodicidad de la intervención.

En el trabajo propuesto por Casey *et al.* (2016) describen el diseño de un robot de compañía (MARIO) para abordar el problema de aislamiento y soledad que suelen padecer las PcDs. Ellos identifican las características que necesita tener un robot de compañía para poder ser aceptado y por lo tanto exista una interacción fluida y frecuente entre el robot y la PcD. Para su estudio reúnen un grupo de enfoque (22 PcDs, 49 enfermeros y 6 familiares) a cuyos participantes realizan una entrevista sobre las características con las que debe contar el robot. Algunas de

las respuestas más frecuentes fueron: que el robot o dispositivo tenga una cara amigable; que tenga la capacidad de hablar a un ritmo lento y con voz fuerte; que les pueda recordar tomar medicamentos y reuniones; conocer la historia de vida del paciente para poder fomentar la conversación sobre eventos que pueden recordar fácilmente y que son de su interés.

En Sarabia *et al.* (2018) desarrollan una interfaz de interacción integrada a un robot humanoide NAO con el objetivo de que pueda ser empleado como robot de compañía en hospitales para mitigar síntomas de depresión y deterioro cognitivo en los pacientes. La interfaz permite que el robot tenga funcionalidades de conversación (empleando el esquema de Mago de Oz), tocar música, bailar y mostrar rutinas de ejercicio. Realizaron una evaluación en un ambiente no controlado con PcDs y con persona sin demencia. Los autores observaron que las PcDs mostraron un mayor interés en tener una nueva sesión con NAO. Sin embargo, durante cada sesión las PcDs mantuvieron un menor tiempo de interacción que las personas sin demencia; también observaron que las PcDs mostraron una clara preferencia en establecer una conversación con NAO en lugar de realizar una actividad física que fuera guiada por el robot.

Actualmente, algunas propuestas ya se enfocan en el desarrollo de herramientas que apoyen a brindar terapia de reminiscencia a pacientes con demencia. Tal es el caso de Sakakibara *et al.* (2017), en donde los autores emplean un agente virtual para realizar las intervenciones con los pacientes. Las intervenciones se realizan a través del uso de plantillas de diálogo que se ajustan a la información de cada paciente. Sin embargo, el diálogo sólo se lleva a cabo en torno al lugar de nacimiento de la persona y no es posible guiar una conversación otros aspectos relevantes de la vida de la PcD como pueden ser: familia, preferencias o estilo de vida.

En la propuesta de Uchiya *et al.* (2018) diseñan un sistema de reminiscencia basado en un agente conversacional por voz empleando la herramienta MMDAgent a través del uso de diversos módulos para la interacción con el usuario como son Julius Speech Recognition Engine, OpenJTalk Voice Synthesis Engine y BulletPyshics. Dentro del sistema emplean un modelo de transición finita para gestionar el diálogo y poder establecer los turnos del sistema y del usuario. Además, establecen un registro de interés en el tema de conversación mediante el número de veces que el usuario habla de ese tema con el sistema. Como técnica de comprensión del lenguaje los autores extraen los sustantivos presentes en la emisión realizada por el usuario y son repetidos por el sistema para ser reafirmados por el usuario, lo

que propicia a que el diálogo se vuelva repetitivo entre el usuario y el sistema. Sin embargo, durante la interacción con el sistema observar que el sistema es capaz de atraer la atención de la PcD y quiera seguir conversando sobre su historia de vida.

3.2.3.1. Trabajos en el idioma español

Analizando trabajos relacionados desarrollados en el idioma español, en la propuesta de Navarro *et al.* (2016) presentan un sistema de cognición asistida para apoyar a cuidadores a brindar terapia ocupacional a PcDs para abordar los síntomas psicológicos y de comportamiento de la demencia. El sistema permite que las PcDs puedan seleccionar alguno de los juegos establecidos (sopa de letras o memorama) y también ofrece recordatorios, información espacio-temporal y ayuda en la realización de actividades. Los autores evaluaron el sistema durante 16 semanas en dos parejas de PcD-cuidador y observaron que la personalización de las intervenciones es un factor relevante para la aceptación y uso del sistema por parte de los pacientes. Sin embargo, al no contar con un SDS integrado dentro del sistema desarrollado un inconveniente es que presenta es que la PcD debe realizar las tareas mediante la interacción con la pantalla o teclado, lo que dificulta su implementación en pacientes que presenten algún problema motriz.

Por otro lado, en el trabajo realizado por Cruz-Sandoval y Favela (2017) describen el diseño de un agente semi-autónomo (Eva), el cual consta de un cuerpo sin movimiento al cual se integra un teléfono inteligente conectado a un dispositivo Raspberry PI. El agente lleva a cabo conversaciones simples con pacientes a través de la coordinación de un operador humano para que pueda ser percibido como un agente conversacional (esquema de Mago de Oz). Eva implementa intervenciones terapéuticas enfocadas a la preservación cognitiva de los pacientes tal como el uso de juegos y terapia de relajación mediante el uso de música. Para determinar la eficacia de Eva, llevan a cabo una evaluación preliminar con un grupo de cuidadores de PcDs (n=8). De los resultados obtenidos concluyen que es importante que el robot pueda ser percibido como un agente autónomo que pueda involucrarse en una interacción con fluidez y la preferencia de comunicación por los pacientes es a través de la voz.

3.3. Conclusión del capítulo

Debido a que la producción y comprensión del lenguaje se mantiene relativamente bien preservada en personas con demencia tipo Alzheimer en una etapa inicial, los sistemas de diálogo por voz o interfaces de voz como herramientas de apoyo en InF han mostrado ser de utilidad y bien adoptadas por las PcDs (Casey *et al.*, 2016; Wolters *et al.*, 2016; Cruz-Sandoval *et al.*, 2018). También se ha observado que al aplicar métricas relacionadas a evaluar la calidad de vida en PcDs, este tipo sistemas ofrecen una mejora en síntomas como depresión y ansiedad. Sin embargo, en la mayoría de las propuestas enfocadas en apoyar a brindar terapia se emplean diálogos genéricos, es decir, el mismo tipo de diálogo es usado para todos los pacientes. Por lo que no proveen los beneficios que se han observado en conversaciones donde el diálogo es personalizado de acuerdo a cada paciente como en la TR. En este sentido, dentro de este trabajo se propone un enfoque para generar diálogos personalizados de acuerdo a la historia de vida y preferencias del usuario para apoyar a brindar la TR de forma constante a personas con demencia tipo Alzheimer en etapa inicial.

Capítulo 4

Metodología

En este capítulo se muestra el diseño, desarrollo e implementación del sistema conversacional resultante. Inicialmente, se describe de forma general la arquitectura del sistema conversacional propuesto. Posteriormente, se describen de forma detallada cada uno de los módulos y elementos que componen el sistema. Finalmente, se delinearán las estrategias de comunicación implementadas en el sistema conversacional para llevar a cabo las sesiones de terapia de reminiscencia.

4.1. Prototipo del Sistema Conversacional

Como parte de los objetivos de este trabajo de investigación se planteó la creación de un prototipo de sistema conversacional que pueda ser utilizado como herramienta auxiliar para brindar terapia de reminiscencia a personas con demencia de tipo Alzheimer en etapa inicial. El diseño del sistema conversacional se basa en la arquitectura clásica de un sistema de diálogo descrita en la sección 2.2. Sin embargo, existen algunas diferencias de acuerdo al dominio de aplicación en el cual se pretende implementar. En la figura 4.1 se muestran los módulos y elementos que componen el prototipo del sistema conversacional propuesto en este trabajo.

El sistema conversacional está integrado de diversos módulos que permiten llevar a cabo una interacción por voz mediante lenguaje natural con el usuario, el cual es una persona con Alzheimer en etapa inicial (PcA). Como interfaz de comunicación entre el sistema y el usuario se utiliza un virtual. Esto debido a que se ha observado

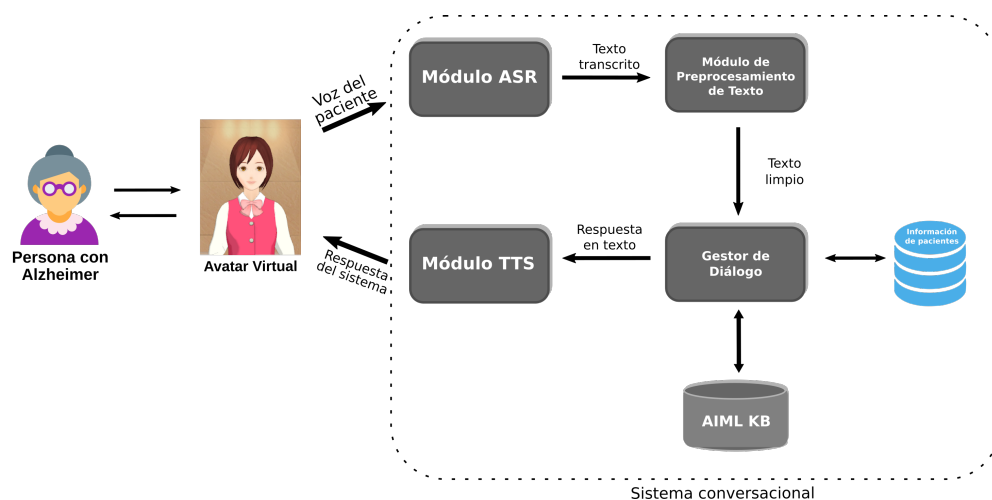


FIGURA 4.1: Arquitectura del sistema conversacional

que pacientes con demencia muestran una respuesta positiva y una mayor adopción e interés al momento de interactuar con sistemas que hacen uso de agentes conversacionales corpóreos (robots sociales, avatares virtuales, humanoides, etc.) contra aquellos que sólo emplean voz o texto para la interacción (Wolters *et al.*, 2016; Casey *et al.*, 2016; Cruz-Sandoval y Favela, 2017; Sarabia *et al.*, 2018). A continuación, se describe de forma general el funcionamiento del sistema, así como las tareas llevadas a cabo por cada uno de los módulos que lo integran. Posteriormente, se detalla la implementación y el funcionamiento de cada uno de los módulos de manera individual.

Como se mencionó anteriormente, la interacción entre el sistema conversacional y la PcA además de ser en lenguaje natural, también se lleva a cabo de forma hablada. Por lo que inicialmente la tarea de procesar el habla del usuario y convertirla a texto se lleva a cabo en el módulo de reconocimiento automático del habla (Módulo ASR), el cual recibe como datos de entrada la voz de la PcA y realiza la tarea de transcribir a texto las emisiones de la PcA, dando como salida el texto transcrito.

Una vez que se cuenta con la emisión transcrita de la PcA, entonces esta es recibida por el Módulo de Preprocesamiento de texto con el objetivo de realizar una limpieza del texto que será procesado por el Gestor de Diálogo. Posteriormente, en el Gestor de Diálogo se analiza la emisión recibida y es precisamente el Gestor de Diálogo quien determina la respuesta que debe de dar el sistema de acuerdo al contexto del diálogo que se este llevando a cabo y a la información de la PcA que esta utilizando el sistema. Para poder realizar esta tarea, el Gestor de Diálogo hace uso de una base de conocimiento creada en AIML y de una base de datos

que contiene la información de las PcAs que interactúan con el sistema. Por un lado, la base de conocimiento permite seleccionar la respuesta más adecuada de acuerdo al contexto del diálogo que se este llevando a cabo y por otro lado, la base de datos permite brindar una respuesta personalizada empleando la información de la PcA que este usando el sistema.

Finalmente, una vez que la respuesta es determinada por el Gestor de Diálogo, esta necesita ser transmitida de forma hablada a la PcA. Por lo tanto, el módulo de conversión de texto a habla (Módulo TTS) se encarga de sintetizar la respuesta y brindarla al usuario.

Hasta este momento se ha mencionado de forma general el funcionamiento del sistema conversacional. Sin embargo, es necesario describir de forma detallada el desarrollo e implementación de cada uno de los elementos que lo integran, los cuales se explican a continuación.

4.1.1. Módulo ASR

Para la implementación de este módulo, se utilizó el servicio web de Google Cloud Speech-to-Text¹ por las características que brinda en cuanto a precisión y tiempo de respuesta en comparación a otras herramientas de reconocimiento del habla disponibles para el idioma español. La implementación de este módulo se realizó de acuerdo al proceso mostrado en la figura 4.2.

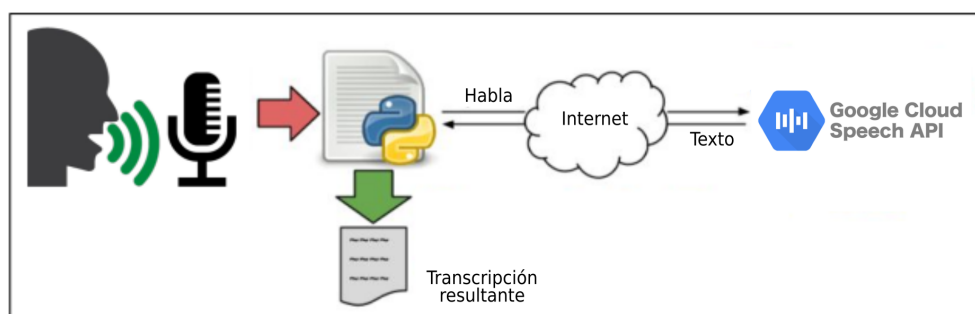


FIGURA 4.2: Implementación del Módulo ASR

Inicialmente, la voz del usuario del sistema conversacional es obtenida empleando un programa realizado en lenguaje Python, dentro de este programa se configuran distintas características del audio capturado, algunas de estas características son:

¹<https://cloud.google.com/text-to-speech/>

el idioma en que se encuentra el audio, la frecuencia del muestreo, el tiempo de espera entre palabras, el formato de audio y el modelo de reconocimiento. Una vez que se ha configurado el formato del audio y las características, se envía una solicitud al servicio de Speech-to-Text, el cual regresa un archivo en formato JSON con las alternativas de transcripción que obtuvieron la probabilidad más alta, como se ilustra en el siguiente ejemplo.

```
{
  "results": [
    {
      {
        "alternatives": [
          {
            "transcript": "Me gusta mucho tocar la guitarra"
            "confidence": 0.9732783748932
          }
        ]
      }
      {
        "alternatives": [
          {
            "transcript": "Me gusta mucho contar la guitarra"
            "confidence": 0.9078937502394
          }
        ]
      }
    }
  ]
}
```

En este caso, el resultado con mayor probabilidad es el que se selecciona del conjunto de alternativas. Una vez que se ha seleccionado la transcripción, esta es pasada al módulo de preprocesamiento de texto como se ilustra en la figura 4.1.

Con el objetivo de utilizar la herramienta de ASR que mostrara el mejor desempeño en el idioma español, se realizó una evaluación comparando el desempeño

de 3 distintas herramientas ASR. Los resultados obtenidos de esta evaluación se describen en la sección 5.1.

4.1.2. Módulo de Preprocesamiento de Texto

Este módulo limpia y normaliza el texto transcrito por el Módulo ASR para eliminar las características del texto que no serán necesarias posteriormente o que incluso pueden afectar al rendimiento del sistema. Las tareas que se llevan a cabo en este módulo son: eliminar los signos de puntuación y los caracteres no alfanuméricos, convertir todos los caracteres a minúsculas y eliminar las palabras repetidas consecutivamente, ya que se ha observado que esta característica se debe a menudo a la deficiencia del reconocedor de voz para hacer frente a los errores de inicio falso en el habla espontánea (p. ej. este este, um um, ee ee). Las tareas realizadas en este módulo permiten posteriormente una mejor comparación entre los enunciados del usuario y los patrones definidos en la base de conocimientos AIML.

4.1.3. Gestor de Diálogo

La parte principal dentro de cualquier sistema de diálogo siempre consiste en como modela y gestiona el diálogo. En este caso, de igual forma el Gestor de Diálogo es el componente que se encarga de las tareas más relevantes dentro del sistema conversacional, como lo son: administrar los turnos entre el usuario y el sistema, seleccionar la respuesta más adecuada de acuerdo al contexto del diálogo, personalizar la respuesta de acuerdo a la información del usuario, así como solucionar problemas durante la interacción.

Si bien existen diferentes tipos de estrategias para el desarrollo de sistemas de diálogo y sistemas conversacionales, actualmente una de las más utilizadas se basa en el uso de técnicas de aprendizaje profundo (Deep Learning) para la creación del modelo y del gestor de diálogo.

La ventaja que ofrecen este tipo de técnicas es que pueden proveer al sistema con la capacidad de aprender de forma implícita la mayoría de estados y reglas para llevar a cabo un diálogo en determinado contexto, lo que permite inicialmente diseñar un conjunto pequeño de reglas y patrones en el modelo de diálogo y a partir de ahí generar los nuevos estados de transición de manera automática. Sin

embargo, una desventaja importante en el uso de estas técnicas radica en el hecho de que requieren de una cantidad considerable de datos para poder desarrollar modelos que realicen un desempeño aceptable (Otter *et al.*, 2020).

En el caso del dominio en el cual se implementa el modelo de diálogo propuesto en este trabajo, se ha constatado que no existe un conjunto de datos (diálogos entre pacientes y cuidadores en donde se brinde terapia de reminiscencia) que permita implementar este tipo de técnicas en el desarrollo del modelado de diálogo para este dominio en particular. Por tal motivo, se empleó un enfoque basado en el reconocimiento de patrones, reglas y estados de información para modelar y gestionar el diálogo. Este tipo de enfoque también permite la creación de agentes conversacionales logrando altas tasas de precisión en la gestión del diálogo para un dominio en específico (Heudin, 2018; Ali *et al.*, 2018).

En el caso del Gestor de Diálogo desarrollado en este trabajo, una de las tareas principales es la personalización del diálogo para cada paciente. Para lograrlo, el gestor necesita interactuar con una base de datos que contiene la información de cada paciente y con una base de conocimientos que se conforma de los contextos de diálogo sobre los cuales el sistema es capaz de conversar.

4.1.3.1. Base de datos de pacientes

Debido a que uno de los principales objetivos es que el sistema genere conversaciones personalizadas acorde a la historia de vida y preferencias de cada paciente, es necesario contar con la información relevante de cada uno de ellos. Para lograr esto, se utilizaron una batería de formularios validados desde la parte médica para determinar que tipo de información es relevante y necesita ser recabada de los pacientes con demencia en centros de cuidado para ofrecer una mejor atención y poder brindar terapia de reminiscencia (Rokkaku, 2005; Phair, 2010).

La información recabada se clasifica en 3 diferentes tópicos (figura 4.3): información personal, familia, historia y estilo de vida del paciente. De acuerdo a los lineamientos establecidos en estos formularios, es preferible que la información sea proporcionada por el cuidador o un familiar del paciente cuando sea posible. De esta batería de formularios se han seleccionado un conjunto de 32 preguntas divididas en los tópicos anteriormente mencionados quedando distribuidas de la siguiente

manera: 11 preguntas referentes a información personal, 4 preguntas referentes a relaciones familiares y 17 preguntas referentes a historia y estilo de vida.

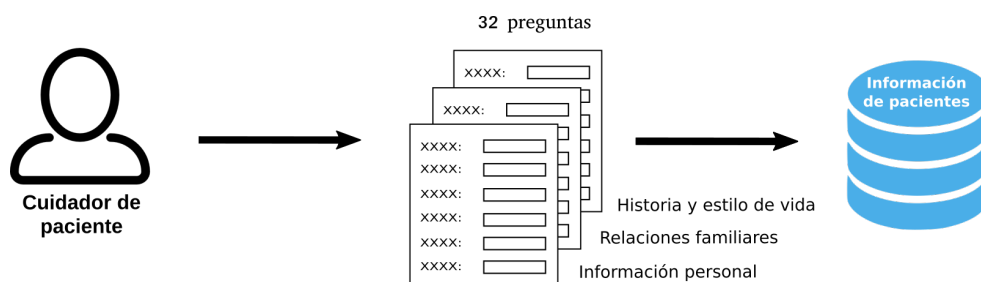


FIGURA 4.3: Recolección de información de pacientes

Para recolectar la información se ha desarrollado una aplicación web que permite ingresar la información a través un formulario dividido en tres secciones. En las figuras 4.4, 4.5 y 4.6 se muestran las preguntas de cada sección.

El formulario, titulado 'Información del paciente', contiene los siguientes campos:

- Nombre completo:
- Como le gusta que le digan: Sexo:
- Fecha de nacimiento: Lugar de nacimiento: Lugar de residencia actual:
- Nivel de estudios: Ocupación: Religión:
- Nombre del cuidador: Parentesco del cuidador:

En la parte inferior del formulario hay un botón azul que dice 'Siguiente'.

FIGURA 4.4: Formulario para el registro de información personal

Los datos recabados a través de estos formularios son requeridos para cada paciente que hará uso del sistema conversacional. Esta información se almacena en una base de datos cuyo modelo se muestra en la figura 4.7. La base de datos es consultada por el gestor de diálogo para crear un perfil para cada paciente y poder personalizar la conversación de acuerdo a la información recolectada de cada uno de ellos.

Información familiar

Nombre del familiar: Parentesco: Edad:

Información adicional

Describir información relacionada con el paciente como: cercanía con el paciente, experiencias vividas juntos, preferencias, hábitos y actividades en común, etc.

Agregar otro

Siguiente

FIGURA 4.5: Formulario para el registro de relaciones familiares

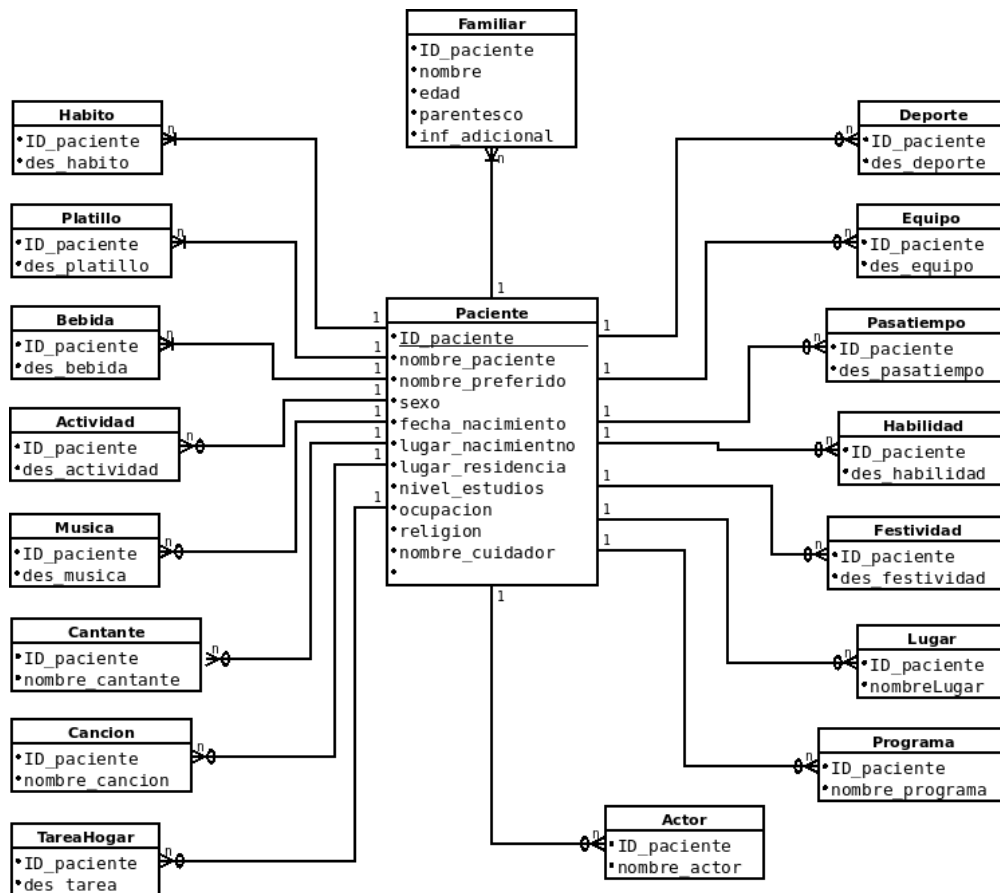
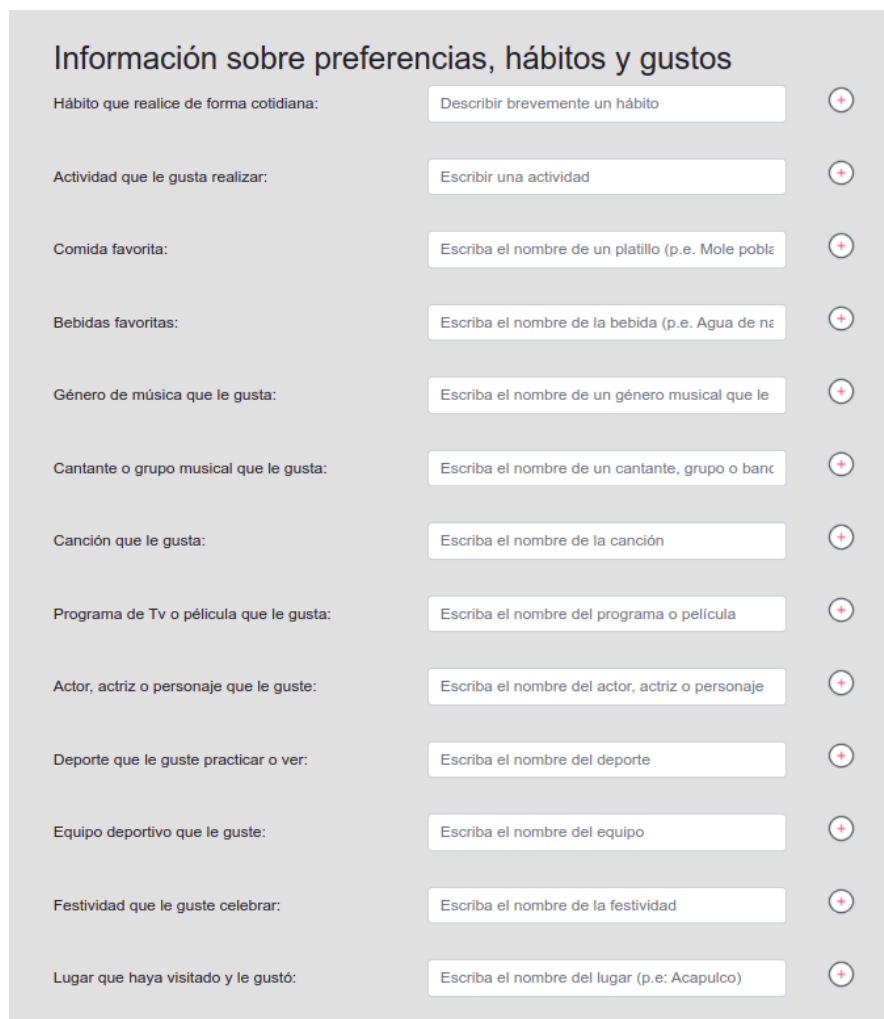


FIGURA 4.7: Modelo de la base de datos Pacientes

En este punto, es importante mencionar que el nivel de personalización de la conversación durante una sesión de terapia esta directamente relacionado a la cantidad de información que se recaba de cada paciente.



El formulario, titulado "Información sobre preferencias, hábitos y gustos", contiene 13 ítems. Cada ítem consiste en una etiqueta de texto a la izquierda, un campo de entrada de texto en el centro y un botón con un signo "+" a la derecha. Los ítems son:

Etiqueta	Placeholder del campo de texto
Hábito que realice de forma cotidiana:	Describir brevemente un hábito
Actividad que le gusta realizar:	Escribir una actividad
Comida favorita:	Escriba el nombre de un platillo (p.e. Mole poblano)
Bebidas favoritas:	Escriba el nombre de la bebida (p.e. Agua de naranja)
Género de música que le gusta:	Escriba el nombre de un género musical que le gusta
Cantante o grupo musical que le gusta:	Escriba el nombre de un cantante, grupo o banda
Canción que le gusta:	Escriba el nombre de la canción
Programa de Tv o película que le gusta:	Escriba el nombre del programa o película
Actor, actriz o personaje que le guste:	Escriba el nombre del actor, actriz o personaje
Deporte que le guste practicar o ver:	Escriba el nombre del deporte
Equipo deportivo que le guste:	Escriba el nombre del equipo
Festividad que le guste celebrar:	Escriba el nombre de la festividad
Lugar que haya visitado y le gustó:	Escriba el nombre del lugar (p.e. Acapulco)

FIGURA 4.6: Formulario para el registro de historia y estilo de vida

4.1.3.2. Modelo de diálogo

Para desarrollar el modelo de diálogo se creó un conjunto de contextos de diálogo empleando el lenguaje AIML para definirlos. La delimitación de estos contextos está basada en los tópicos sobre los que el sistema conversacional es capaz de dialogar con el usuario, los cuales de acuerdo a lo descrito en la sección 4.1.3.1 son: familia, historia y estilo de vida. Estos contextos de diálogo son los que se describen en la tabla 4.1.

Cada contexto se representa mediante un objeto AIML en donde se definen los posibles patrones de entrada y sus respuestas en ese contexto. En la figura 4.8 se muestra la forma más simple de definir un contexto, el cual en su estructura básica debe de contar por lo menos con una categoría (*category*) que este conformada por un patrón (*pattern*) y al menos un tipo de respuesta para dicho patrón (*template*).

Contexto de diálogo	Descripción (Conversaciones sobre)
Conversación inicial	Saludo de bienvenida e inicio de la sesión
Finalización de la sesión	Despedida del paciente
Información del agente	Información sobre el agente conversacional (p.e., nombre, genero, edad, etc.)
Perfil del paciente	Información personal del paciente (p.e., cumpleaños, apodo, lugar de nacimiento)
Familia	relaciones familiares relevantes para el paciente
Habitos	Hábitos que el paciente tiene
Habilidades	Habilidades que el paciente tiene (p.e., tejer, dibujar, cantar, carpintería)
Pasatiempos	Pasatiempos que el paciente disfruta realizar
Tareas domésticas	Tareas en las que el paciente le gusta ayudar
Lugares visitados	Lugares relevantes en la vida del paciente
Preferencias del paciente sobre	
Películas	Películas que le gustan
Programas de televisión	Programas de su época favoritos
Actores	Actores o personajes de interés
Música	Su tipo de música que es preferida
Cantantes	Cantantes o grupos musicales que le gustan
Deportes	Deportes que son de agrado del paciente
Equipos deportivos	Equipos deportivos o deportistas
Platillos de comida	Platillos de comida que le gustan
Bebidas	Bebidas que le gustan
Festividades	Festividades de las cuales tiene buenos recuerdos (p.e., navidad, día de las madres, cumpleaños)

TABLA 4.1: Contextos de diálogo definidos. Los contextos descritos son sobre los cuales el sistema tiene la capacidad de llevar a cabo la terapia.

La creación de diversos contextos permite ajustar mejor el flujo de diálogo entre el usuario y el sistema para cada contexto en particular. El proceso llevado a cabo para el modelado del diálogo es el siguiente:

1. Se recibe la oración de entrada por parte del Módulo de Preprocesamiento de Texto.
2. Dentro del conjunto de contextos definidos el modelo de diálogo busca un patrón que coincida con la oración recibida. Si se encuentra un patrón que coincida con la oración en alguno de los contextos entonces se obtienen las posibles respuestas asociadas a dicho patrón.
3. Por último, el gestor de diálogo determina que respuesta es seleccionada y procesada por el módulo TTS para ser entregada al usuario.

Además, se han considerado diferentes casos que podrían presentarse al interactuar con el usuario y que el gestor de diálogo debe tratar de solucionar, como lo son


```

<?xml version="1.0" encoding="UTF-8"?>

<!-- base de conocimiento para el contexto de "presentacion" -->
<aiml version = "2.0">
  <category>
    <pattern>HOLA</pattern>
    <template>
      <random>
        <li> ¡Hola! ¿Cómo estas?</li>
        <li> ¡Qué tal! ¿Cómo estas?</li>
        <li> Me da gusto verte de nuevo. ¿Cómo estas?</li>
        <li> Me da gusto saludarte. ¿Cómo estas? </li>
      </random>
    </template>
  </category>

  <category>
    <pattern>COMO TE LLAMAS</pattern>
    <template>
      <random>
        <li> Mi nombre es <bot name="name"></bot></li>
        <li> Soy <bot name="nombre"></bot> y me gustaría mucho platicar contigo</li>
        <li> Mis amigos me dicen <bot name="nombre"></bot></li>
      </random>
    </template>
  </category>
</aiml>

```

FIGURA 4.8: Ejemplo de la definición en AIML del contexto *Conversación inicial*

la asociación de patrones diferentes que tienen la misma interpretación semántica. Por ejemplo, la oración de entrada “*Recuérdame alguna canción que cantaba X*”, donde X representa una entidad (persona), podría tener la misma intención que la sentencia “*¿Qué canciones cantaba X?*”. Para estos casos se ha implementado la técnica de Reducción Simbólica con el objetivo de realizar una simplificación de los patrones. Por lo que patrones de una gramática compleja pueden ser trasladados a patrones más simples que pueden retornar la misma respuesta. La figura 4.9 ilustra este procedimiento.

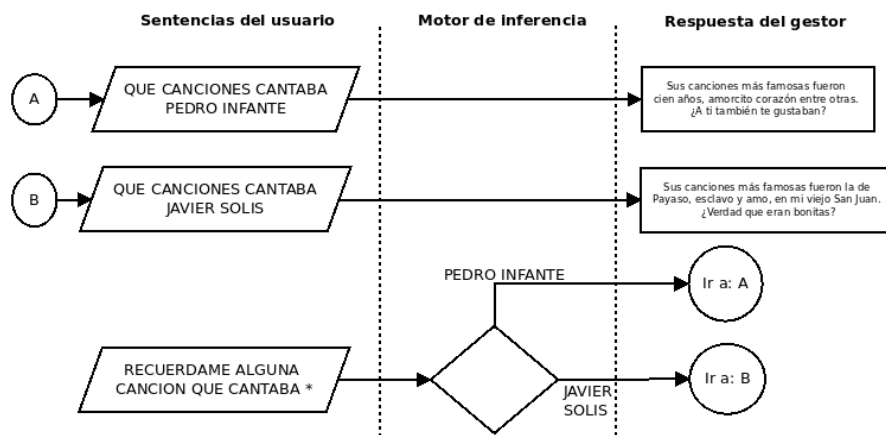


FIGURA 4.9: Reducción simbólica aplicada en el contexto de “cantante”

Lo que muestra la figura es que el gestor de diálogo puede recibir distintos patrones por parte del usuario, sin embargo, cuando los patrones recibidos sean el A y el B, entonces el gestor asigna una respuesta directa a esos patrones. Sin embargo, cuando el patrón es “*Recuérdame alguna canción que cantaba **”, entonces se toman las palabras representadas por * (Pedro Infante o Javier Solís en este ejemplo) y se relaciona a alguno de los dos patrones establecidos A o B, dependiendo del valor de *. Este proceso de reducción permite que aunque existan una gran variedad de patrones con la misma intención o solicitud, estos puedan ser asociados a la respuesta existente de un solo patrón con el objetivo de simplificar las transiciones entre los distintos estados del diálogo.

4.1.3.3. Personalización del diálogo

Como se ha mencionado a lo largo del documento, uno de los principales objetivos es la personalización del diálogo para que el sistema conversacional pueda ofrecer los beneficios de la terapia de reminiscencia. En este sentido, la personalización del diálogo que el sistema ofrece al usuario, se basa en la información recabada y almacenada en la base de datos descrita en la sección 4.1.3.1 para utilizarla en un conjunto de plantillas de respuesta creadas dentro de los diferentes contextos de diálogo establecidos. En la figura 4.10 se muestran dos ejemplos de las plantillas creadas para integrar la información del paciente en los diálogos de acuerdo al contexto.

```
<topic name="saludo">
  <category>
    <pattern>HOLA</pattern>
    <template>
      <random>
        <li> ¡Hola! ¿Cómo estas <set name="nombrePaciente"> $User.nombre</set>?</li>
        <li> ¡Qué tal <set name="nombrePaciente"> $User.nombre</set>! ¿Cómo estas?</li>
        <li> Me da gusto verte de nuevo <set name="nombrePaciente"> $User.nombre</set>. ¿Cómo estas?</li>
      </random>
    </template>
  </category>
</topic>

<topic name="familia"
  <category>
    <pattern>_ FAMILIA *</pattern>
    <template>
      <random>
        <li> Es verdad que <set name="nombreFamiliar1"> $User.familiar1.nombre</set> es tu <set name="parentescoFamiliar1"> $User.familiar1.parentesco</set> </li>
        <li> Te gustaría platicar de tu <set name="parentescoFamiliar1"> $User.familiar1.parentesco</set> <set name="nombreFamiliar1"> $User.familiar1.nombre</set>
      </random>
    </template>
  </category>
</topic>
```

FIGURA 4.10: Plantillas personalizadas con información del paciente

En el primer ejemplo (parte superior), se puede observar que el contexto del diálogo es el de saludar (<topic name="saludo">) y que en las plantillas generadas se emplea

la información recuperada de la base de datos referente únicamente al nombre del paciente en este caso. Por lo que las posibles repuestas que se darían al patrón HOLA si `$User.nombre=Carlos` serían: *‘¡Hola! ¿Cómo estas Carlos?’*, *“¡Qué tal Carlos! ¿Cómo estas?”* o *“Me da gusto verte de nuevo Carlos. ¿Cómo estas?”*.

En el segundo ejemplo, el contexto del diálogo es “familia” y la información que se utiliza es el nombre del primer familiar registrado en la base de datos y su parentesco con el paciente. De tal forma que si `$User.familiar1.nombre=Patricia` y `$User.familiar1.parentesco=hija`, entonces las posibles respuestas al patrón `<pattern>_ FAMILIA *<pattern>` de acuerdo a la plantilla creada serían: *“Es verdad que Patricia es tu hija”* o *“Te gustaría platicar de tu hija Patricia”*.

En los ejemplos anteriores únicamente se emplea determinada información; sin embargo, toda la información relacionada al paciente es recuperada de la base de datos y almacenada en la memoria del gestor de diálogo debido a que existen diferentes plantillas en las que se emplea la demás información.

4.1.4. Módulo TTS

El Módulo TTS o módulo de texto a habla realiza la tarea de sintetizar la respuesta textual generada por el Gestor de Diálogo a una forma hablada. El servicio utilizado en este módulo fue el de Google Cloud Text-to-Speech². Implementado las estrategias de comunicación recomendadas para comunicarse verbalmente con las PcAs, las cuales se describen en la sección 4.2.2, se realizaron algunos ajustes en la síntesis del texto a habla. Algunos de estos ajustes fueron: reducir la velocidad del habla a un valor de 0.7 con respecto a la velocidad normal del habla, con el objetivo de que la respuesta del sistema pueda ser percibida de forma clara y lenta por la PcA; emplear un tono de voz femenina, ya que se ha observado que es preferida por las PcA;

También se configuraron valores referentes a la cadencia y entonación al sintetizar palabras específicas como las clasificadas como saludo, despedida, pregunta, solo por mencionar algunas. Esto con el objetivo de que la respuesta sea percibida de forma natural por el PcA.

El proceso llevado a cabo por este módulo es bastante similar al del Módulo SST, pero de forma invertida, ya que en este caso primero se recibe la respuesta del

²<https://cloud.google.com/text-to-speech>

Gestor, posteriormente es enviada al servicio de TTS y el audio recibido es dado como respuesta al usuario como se puede observar en la figura 4.11.

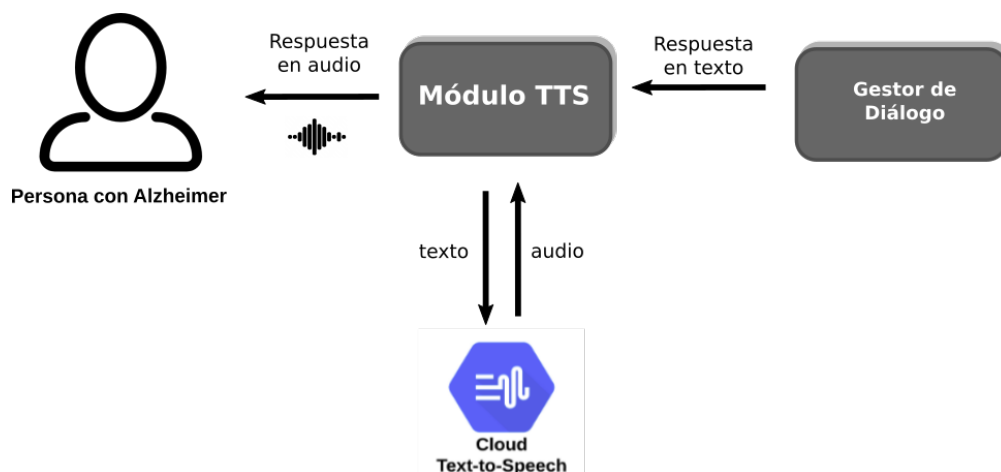


FIGURA 4.11: Implementación del Módulo TTS

Es importante mencionar que tanto el servicio de Speech-to-Text como el de Text-to-Speech son servicios basados en la nube, por lo que para que puedan ser utilizados es necesario que el dispositivo en el cual se utilice el sistema conversacional cuente con una conexión a Internet.

4.2. Metodología para llevar a cabo la terapia de reminiscencia por el sistema conversacional

Las condiciones que se consideran dentro del proceso de brindar la terapia y que se describen a continuación son las siguientes: las características que deben tener los pacientes que podrán interactuar con el sistema, el diseño de las intervenciones, los instrumentos y diseño de evaluación y por último las cuestiones éticas a considerar.

4.2.1. Características de los participantes

Es importante recalcar que el uso del prototipo del sistema conversacional como se mencionó en los objetivos de este trabajo, está limitado a ser usado por PcA en

etapa inicial debido a las características degenerativas progresivas que se manifiestan en este padecimiento. Por lo que algunos criterios de exclusión para interactuar con el sistema son:

- Que la PcA cuente con algún diagnóstico psiquiátrico adicional.
- Que obtengan una puntuación en el MMSE < 15 o que haya sido diagnosticado como PcA en etapa intermedia-avanzada.
- Que la PcA padezca una enfermedad física grave, discapacidad sensorial o alto nivel de agitación.
- Que tenga algún trastorno de la visión o la audición que los incapacitara para poder interactuar con el sistema.

Además, resulta necesario describir las principales alteraciones que presenta una PcA en una etapa inicial. La tabla 4.2 describe algunas de estas alteraciones.

Alteración o síntoma	Descripción
Perdida de eventos recientes	Las personas podrían tener dificultades para recordar la información recién obtenida y hacer la misma pregunta una y otra vez.
Dificultad para resolver problemas, tareas complejas y juicios sólidos	Planificar un evento cotidiano o llevar a cabo su administración financiera puede resultar un proceso difícil.
Cambios de personalidad	Las personas pueden sentirse sometidas o retraídas, especialmente en situaciones socialmente difíciles, o mostrar una irritabilidad o un enojo inusuales. También es común la reducción de la motivación para completar las tareas.
Dificultad de comunicación	Los primeros problemas de comunicación se presentan con casos crecientes de anomia provocados por el deterioro de la memoria semántica
Perderser o extraviar pertenencias	Las personas tienen ocasionalmente dificultades para orientarse, incluso en lugares conocidos. Comienza a ser común extraviar cosas.
Nivel de comprensión	La comprensión aun se encuentra bastante preservada en esta fase
Habilidades pragmáticas	Comienza a observarse una disminución de la conducta conversacional iniciativa; la persona se fija una posición en su mayoría de solo respuesta.

TABLA 4.2: Descripción de síntomas en PcA en etapa inicial

Se tiene que tener en cuenta que aunque existe una descripción general de las alteraciones cognitivas o sintomatología que puede presentar una PcA en etapa inicial, el desarrollo y aparición de estas alteraciones puede variar significativamente de un paciente a otro en cuanto a intensidad y duración. En este sentido, el sistema desarrollado no tiene como objetivo adaptarse a la severidad presentada en cada etapa del Alzheimer, por lo que se limita su uso a PcA que cuenten con las características mencionadas anteriormente.

4.2.2. Diseño de las intervenciones

En diversos trabajos se han propuesto estrategias para establecer y mejorar la comunicación verbal con PcA (Small y Gutman, 2002; Gonzalez *et al.*, 2015; Hernández-Domínguez *et al.*, 2016). Algunas de las estrategias más relevantes recomendadas en común son las se describen a continuación:

- Utilizar frases cortas y simples.
- Hablar claro y lento.
- Preguntar o dar una instrucción a la vez.
- Evitar interrumpir al paciente.
- Utilizar en lo posible preguntas cerradas (si/no) en lugar de preguntas abiertas.
- De ser necesario, repetir el mensaje utilizando las mismas palabras.
- Conversar acerca de los temas de interés del paciente.
- Personalizar la conversación.
- Utilizar elogios y frases de refuerzo.
- No forzar a hablar al paciente.
- Identificar cuando un paciente quiere terminar la conversación

Estas estrategias deben de ser consideradas cuando se lleva a cabo una comunicación con PcA debido al deterioro cognitivo que conlleva este padecimiento. Por lo que, que estas estrategias fueron consideradas para su integración en el sistema conversacional. Un vez que se han establecido las estrategias a seguir por el sistema, se determina la forma en que el sistema va a interactuar con el paciente.

Inicialmente, el contexto del diálogo o tema de conversación entre el paciente y el sistema es determinado por el cuidador del paciente basado en el conocimiento que tiene el cuidador sobre los temas que le gusta conversar y que estos temas se encuentren dentro de los contextos determinados en el sistema conversacional. Además, es el cuidador del paciente el encargado de explicarle como es que se lleva a cabo la interacción con el sistema.

Capítulo 5

Experimentos y resultados

En este capítulo se describen por una parte los experimentos realizados para comparar los sistemas de reconocimiento automático del habla con el objetivo de seleccionar el que ofreciera el mejor desempeño y por otro lado se muestra la evaluación del sistema conversacional para llevar a cabo la terapia de reminiscencia.

5.1. Comparación evaluativa de sistemas ASR

Debido a que como parte del desarrollo del sistema conversacional se implementa un módulo para el reconocimiento automático del habla (Módulo ASR), es importante determinar que sistema ASR ofrece un mejor desempeño en esta tarea, ya que el error cometido por el módulo ASR en el reconocimiento afecta de forma directa el desempeño del sistema conversacional.

En este sentido, se realizó una evaluación comparativa de distintos sistemas de ASR comerciales, tomando en cuenta aquellos que mencionan tener una precisión arriba del 85 % de precisión en el reconocimiento del idioma español. Para realizar la evaluación se recolectaron y emplearon archivos de audio que contienen conversaciones con personas de la tercera edad sin ningún tipo de deterioro cognitivo diagnosticado. Esto con el fin de observar el desempeño de estos sistemas ante situaciones que se pueden presentar en este grupo poblacional como pueden ser: 1) pausas largas al pronunciar la siguiente palabras u oración, 2) un mayor número de inicios falsos, 3) variaciones en la entonación, por mencionar algunos.

Los sistemas ASR analizados dentro de este trabajo fueron: Google Speech-to-Text¹ (**ASR-1**), IBM Watson Speech to Text² (**ASR-2**) y Amazon Transcribe³ (**ASR-3**). Cada uno de los sistemas anteriores tiene características distintas que de acuerdo a sus afirmaciones, ofrecen un mejor desempeño en determinados casos de uso. Sin embargo, como se mencionó anteriormente para este contexto en específico en donde se implementó un módulo de ASR dentro de un sistema conversacional enfocado a PcA, que en su mayoría son personas de la tercera edad, es necesario evaluar el desempeño de los sistemas ASR dentro de este escenario. Por lo que a continuación se describen los experimentos realizados y los resultados obtenidos al evaluar estos sistemas.

5.1.1. Conjunto de datos

Los archivos de audio para evaluar los sistemas ASR fueron recabados a partir de un cuestionario realizado a 8 adultos mayores (> 65 años). Cada cuestionario fue grabado en audio y su uso en esta investigación cuenta con el consentimiento de cada participante. Las preguntas planteadas en el cuestionario fueron las siguientes:

1. ¿En qué año nació?
2. ¿En dónde nació?
3. Describa brevemente algo que le guste del lugar donde nació
4. ¿Puede mencionar algún recuerdo agradable de su niñez?
5. Describa brevemente su rutina en un día normal

En la tabla 5.1 se muestran las características de los archivos de audio de acuerdo a cada participante.

Como se puede observar la mayoría de los participantes son mujeres ($n = 5$) y la edad de los participantes se encuentra en un rango entre 65 y 81 años; la duración total de los archivos de audio recolectados es de 36.37 minutos. De cada archivo de audio fueron eliminadas las preguntas (realizadas por el entrevistador), dejando

¹<https://cloud.google.com/speech-to-text>

²<https://www.ibm.com/cloud/watson-speech-to-text>

³<https://aws.amazon.com/transcribe/>

TABLA 5.1: Descripción de archivos de audio

Archivo generado	Sexo	Edad	Duración (minutos)
AP1	Femenino	67	3.28
AP2	Masculino	68	5.32
AP3	Femenino	65	4.15
AP4	Femenino	81	4.10
AP5	Masculino	78	6.18
AP6	Femenino	76	3.14
AP7	Femenino	67	5.50
AP8	Masculino	70	4.30

únicamente las respuestas brindadas por cada participante, esto con el objetivo de que la voz del encuestador no modifique los resultados obtenidos por cada sistema.

Es importante mencionar que todos los participantes en las entrevistas son personas que aun cuentan con sus habilidades comunicativas bien preservadas y sin ninguna enfermedad detectada que afecte su capacidad cognitiva y de lenguaje.

5.1.2. Experimentos

Para evaluar la precisión de la transcripción de un sistema ASR una métrica comúnmente usada es la tasa de error por palabra o Word Error Rate (WER) (Ramírez~Sánchez *et al.*, 2019; Park *et al.*, 2008), el valor WER es un valor numérico que se calcula mediante la siguiente formula:

$$WER = \frac{S + I + D}{N} \quad (5.1)$$

donde:

- S representa el número de sustituciones, es decir, cada vez que una palabra es reemplazada (por ejemplo: “siempre queríamos salir” es transcrita como “siempre queríamos abrir”).
- I representa el número de inserciones, es decir, cada vez que se agrega una palabra que no se dijo (por ejemplo: “comía temprano” es transcrita como “comía de temprano”).

- D representa el número de eliminaciones, es decir, cada vez que una palabra es omitida o eliminada (por ejemplo: “caminaba con mi perro” es transcrita como “caminaba mi perro”).
- N representa el número total de palabras que fueron transcritas.

Como se puede observar, por lo tanto el valor WER es el número de veces que el sistema ASR substituyó, insertó o eliminó palabras del texto de referencia entre el número total de palabras transcritas. En este sentido, un valor de $WER = 0$ indicaría una transcripción perfecta, mientras que un valor elevado indica una baja precisión por parte del sistema ASR.

Para evaluar los sistemas ASR inicialmente se realizó la transcripción manual de cada archivo de audio, las cuales son tomados como *transcripción de referencia* para comparar la transcripción brindada (transcripción hipótesis) por cada sistema. Dentro de la transcripción manual se conservaron palabras que pudieran ser casos de disfluencia que comúnmente se presentan al hablar considerando que al ser pronunciadas dentro del habla, el sistema ASR debería de ser capaz de reconocerlas.

Posteriormente cada archivo de audio fue procesado por cada uno de los tres sistemas ASR y la transcripción hipótesis generada por cada sistema fue comparada con la transcripción de referencia y de esta forma se calculó el valor WER para cada archivo analizado.

TABLA 5.2: Valores WER obtenidos para cada archivo de audio

Archivo	ASR-1	ASR-2	ASR-3
AP1	0.28	0.34	0.39
AP2	0.19	0.23	0.27
AP3	0.31	0.40	0.43
AP4	0.22	0.21	0.34
AP5	0.25	0.36	0.50
AP6	0.29	0.32	0.37
AP7	0.24	0.39	0.41
AP8	0.37	0.41	0.45
Promedio	0.268	0.332	0.395

5.1.3. Resultados

La tabla 5.2 muestra los valores WER obtenidos para cada archivo, así como el promedio del valor WER para cada sistema. Se puede observar que el sistema de Google Speech-to-Text (ASR-1) muestra una mejor precisión al tener el menor valor promedio de WER (0.268). En la figura 5.1, también se puede observar que el desempeño del sistema ASR-1 mostró un comportamiento más estable durante el procesamiento de todos los archivos de audio.

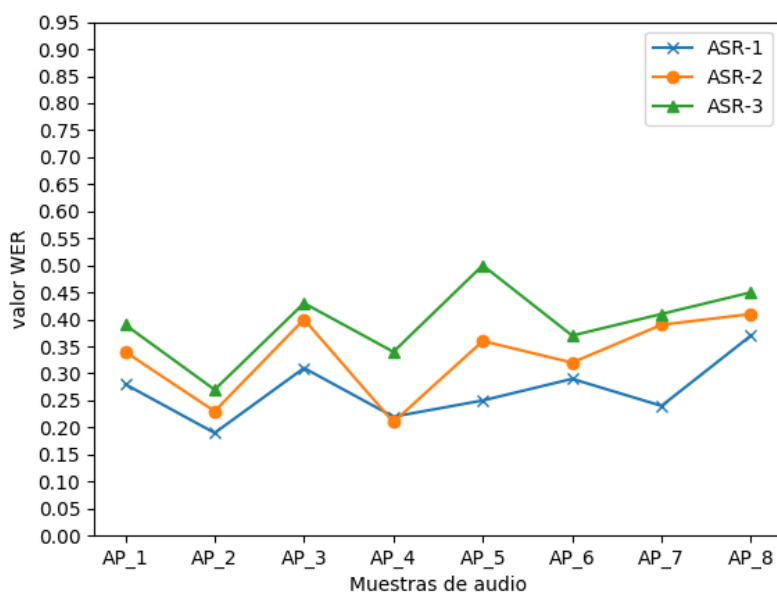


FIGURA 5.1: Comparación de los valores WER obtenidos

Algunas observaciones adicionales obtenidas durante la realización del experimento fueron que: 1) el sistema ASR-1 mostró un mejor desempeño al detectar sustantivos propios (nombres y ciudades), 2) el sistema ASR-1 reconoció de forma más precisa números y cantidades monetarias y 3) El sistema ASR-1 mostró un mejor desempeño al reconocer platillos de comida. Por lo que de acuerdo a los resultados obtenidos se puede determinar que el sistema de Google Speech-to-Text obtiene la mejor precisión en comparación con los otros dos sistemas al menos en este escenario en particular, en el cual el reconocimiento del habla es de personas de la tercera edad.

5.2. Evaluación del sistema conversacional

La evaluación del sistema conversacional se enfocó en este momento en determinar que tan apropiadas son las respuestas que brinda el sistema para llevar a cabo la terapia de reminiscencia en PcA. En este sentido, el sistema fue evaluado por cuidadores, familiares o personas relacionadas al cuidado de PcA.

Para llevar a cabo esta evaluación se siguió un esquema de evaluación estático (Gandhe y Traum, 2016) con el objetivo de evaluar la idoneidad de las respuestas dadas por el sistema a las emisiones del usuario en un contexto de diálogo determinado. Este esquema propone que, para facilitar la participación de los evaluadores, éstos adopten el papel de espectadores. Esto significa que el evaluador no participa activamente en la conversación y así más de un evaluador puede evaluar la misma conversación, con lo que se logra una disminución en los tiempos y proceso de evaluación.

Por lo tanto, el evaluador no interactúa directamente con el sistema pero se le brinda un conjunto de conversaciones llevadas a cabo entre el sistema y un usuario. Cada una de estas conversaciones esta compuesta por un contexto de diálogo (de al menos 4 turnos de interacción entre el sistema y el usuario) y un conjunto de respuestas candidatas como se ilustra en la figura 5.2, de las respuestas candidatas una de ellas es la que brinda el sistema. Para cada una de las conversaciones evaluadas, el evaluador selecciona la respuesta que a su consideración es la que debería de dar el sistema de acuerdo al contexto de diálogo dado en cada conversación.

5.2.1. Configuración de la evaluación

Para llevar a cabo la evaluación se generó un conjunto de 40 conversaciones entre el sistema y un usuario (paciente simulado). La información del paciente empleada para la personalización de la terapia fue generada ya que hasta el momento no se ha tomado de algún paciente real. Para generar el conjunto de conversaciones que se evaluó, se tomaron en cuenta todos los contextos de diálogo sobre los cuales el sistema tiene la capacidad de conversar y que fueron descritos en la sección 4.1.3.2, por lo que se generaron 2 conversaciones por contexto.

La forma de poner a disposición de los evaluadores el conjunto de conversaciones fue mediante un formulario web. La figura 5.2 muestra una captura de un ejemplo

del conjunto de conversaciones, las cuales como ya se mencionó están compuestas por un contexto de diálogo y las respuestas candidatas.

5 → El contexto de la conversación es : saludo

Sistema: *Que tal don Amador. ¿cómo se encuentra el día de hoy?*

Paciente: *Un poco cansado*

Sistema: *Me apena oír eso. Pero, ¿qué le parece si platicamos un poco para que se anime?*

Paciente: *Si*

Sistema:

¿Cuál de las siguientes oraciones consideras que sería la respuesta más adecuada que tendría que dar el sistema de acuerdo a la conversación anterior?

Tecla **A** - Perfecto, vera que después de platicar un rato se siente mejor

B ¿De qué le gustaria platicar?

C Muy bien, ¿le parece si platicamos un poquito sobre la familia?

D Así se habla, verá que vamos a pasar un rato muy agradable platicando

FIGURA 5.2: Ejemplo de una conversación donde el evaluador tiene que seleccionar la respuesta más apropiada de acuerdo al contexto de diálogo

En el ejemplo mostrado en la figura 5.2, el evaluador tiene que votar (seleccionar) por la respuesta que a su consideración es la más apropiada respecto al contexto de diálogo que se esta llevando a cabo. Este proceso fue realizado por cada evaluador para cada una de las 40 conversaciones del conjunto de evaluación.

El método de reclutamiento de los evaluadores fue a través de una convocatoria de participación publicada en distintos grupos de redes sociales relacionados con brindar información y apoyo a los familiares o cuidadores de PcAs. De las personas que manifestaron su interés en participar, se seleccionaron aquellas que cumplían los siguientes criterios de selección: (1) tener una PwAD bajo su cuidado directo y pertenecer a una de las siguientes categorías: terapeuta, médico geriatra, auxiliar de centro de atención o cuidador principal, y (2) estar familiarizado con la terapia de reminiscencia y haberla impartido al menos una vez.

De un total de 18 personas que expresaron su interés en participar, se seleccionaron 11 personas que cumplían los criterios de selección. Por lo tanto los 11 evaluadores seleccionados pertenecen a las siguientes categorías: cuidadores primarios (n=9), médico geriatra (n=1), personal de asilo (n=1). Es importante resaltar que ninguno

de los participantes recibió algún tipo de compensación económica por participar en este estudio. El conjunto de conversaciones evaluadas se encuentra disponible en Internet (https://github.com/vicmman/cs_evaluation)

Una vez concluido el proceso de evaluación, se obtiene una puntuación de los votos obtenidos de cada respuesta dentro de cada conversación del conjunto. Entonces, se realizó la evaluación del sistema tomando como base la métrica de evaluación denominada *Voted Appropriateness* propuesta por Gandhe y Traum (2016). Se calcula el valor $R_{voted} = V(u_t, context_t)$ para cada respuesta. El valor R_{voted} representa el número de evaluadores que seleccionaron la respuesta (*utterance*) u_t como la respuesta más apropiada de acuerdo al contexto de diálogo $context_t$ en la conversación evaluada t . Los resultados de la evaluación se detallan a continuación.

5.2.2. Resultados

Como se mencionó anteriormente cada uno de los diálogos del conjunto fue evaluador por los 11 evaluadores. De igual forma, para cada respuesta se calculó su valor R_{voted} . Por lo que, de acuerdo al número de evaluadores el valor R_{voted} máximo para cada respuesta puede ser de 11 si la respuesta es seleccionada como apropiada por todos los evaluadores y el valor mínimo sería de 0 si la respuesta no fue seleccionada como apropiada por ningún evaluador.

En este sentido, en la figura 5.3 se muestran los valores R_{voted} obtenidos por cada respuesta en cada uno de las conversaciones del conjunto evaluado. Se puede observar que los valores $R_{voted}(sys_res_t, context_t)$ correspondientes a las respuestas dadas por el sistema son más altos que los otros valores R_{voted} en la mayoría de las conversaciones evaluadas, únicamente seguidos por los valores $R_{voted}(can_res1_t, context_t)$ correspondientes a la respuesta candidata 1. Los valores R_{voted} correspondientes a las respuestas candidatas 2 y 3 revelan que estas fueron las menos votadas, obteniendo valores máximos de 4 y 2 votos en pocos diálogos respectivamente.

Estos resultados permiten aseverar que las respuestas dadas por el sistema conversacional dentro de cada diálogo fueron en la mayoría de los casos las respuestas más apropiadas de acuerdo al número de votos recibidos por los evaluadores.

Además, una vez que se calcularon los valores R_{voted} para cada respuesta, también se obtiene la correlación lineal entre el número de votos obtenidos por las respuestas

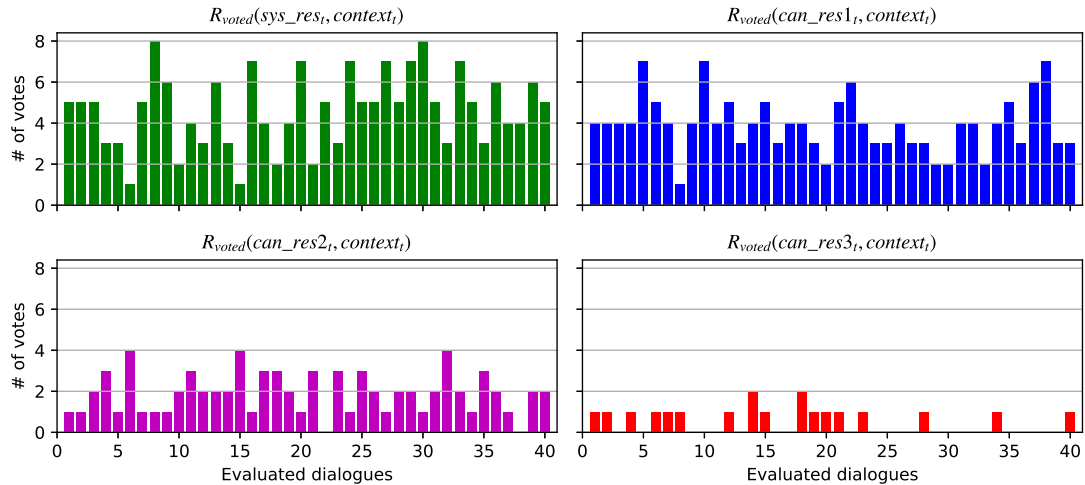


FIGURA 5.3: Valores R_{voted} obtenidos por cada respuesta en el conjunto de diálogos evaluados

del sistema $R_{voted}(sys_res_t, context_t)$ y las respuestas que obtuvieron el máximo número de votos en cada conversación evaluada, representadas por el valor R_{max} como se muestra en la figura 5.4. Se calculó un coeficiente de correlación de Pearson de 0.634 ($p < 0.001, n = 40$), con lo que se puede afirmar que existe una correlación positiva entre estas variables. Y efectivamente, se puede observar a lo largo de la evaluación que existe una correlación en que la respuesta dada por el sistema es una de las más votadas en cada conversación del conjunto evaluado.

Siguiendo esta línea, en la figura 5.5 se puede observar que las respuestas que brinda el sistema obtienen el más alto porcentaje (62.5 %) como las respuestas más votadas por los evaluadores, mientras que las respuestas candidatas 1 obtienen el 42.5 %, las respuestas candidatas 2 solo el 2.5 % y las respuestas candidatas 3 no coinciden en ninguna ocasión como la respuesta más votada.

Los resultados obtenidos siguiendo el esquema de evaluación descrito anteriormente muestran que aunque el número máximo de votos que puede alcanzar una respuesta en una conversación del conjunto de evaluación es de 11 votos, el mayor número de votos alcanzado por una respuesta fue de 8 votos, coincidiendo con los votos obtenidos por la respuesta del sistema como se observa en la figura 5.4. Que todos los evaluadores coincidan en que una única respuesta es la apropiada resulta difícil, ya que existe un grado de subjetividad durante el proceso de evaluación, ya que mientras para una persona puede ser una respuesta adecuada para otra persona puede no serlo y en este caso las dos podrían estar en lo correcto.

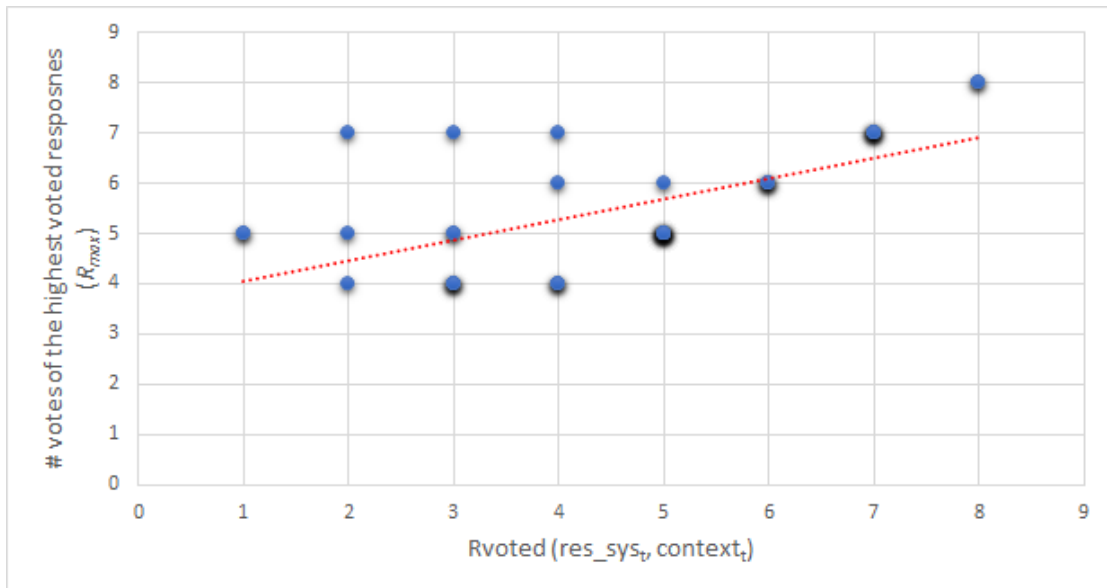


FIGURA 5.4: Correlación entre los valores $R_{voted}(res_sys_t, context_t)$ y R_{max} . Una pequeña cantidad de sombra fue añadida a los valores R_{voted} para visualizar la mayoría de los casos

Aunque la mejor forma de evaluar el desempeño de un sistema conversacional sigue siendo el juicio humano, hay muchas ocasiones en las que las respuestas generadas por el sistema no suelen contar con el acuerdo de todos los evaluadores. Por lo que, un buen indicador de un buen desempeño del sistema suele ser que la mayoría de los evaluadores estén de acuerdo en que las respuestas del sistema son las más adecuadas según el contexto del diálogo, como en este caso. De acuerdo a esto, los resultados de la evaluación permiten afirmar que las respuestas dadas por el sistema en los distintos escenarios de diálogo son las más apropiadas en función del número de votos otorgados por los evaluadores a cada respuesta, alcanzando el mayor número de votos en la mayoría de las conversaciones del conjunto de evaluación.

De esta forma, se puede considerar que el sistema conversacional desarrollado es adecuado para poder ser usado como herramienta auxiliar en sesiones de terapia de reminiscencia con pacientes de Alzheimer en etapa inicial.

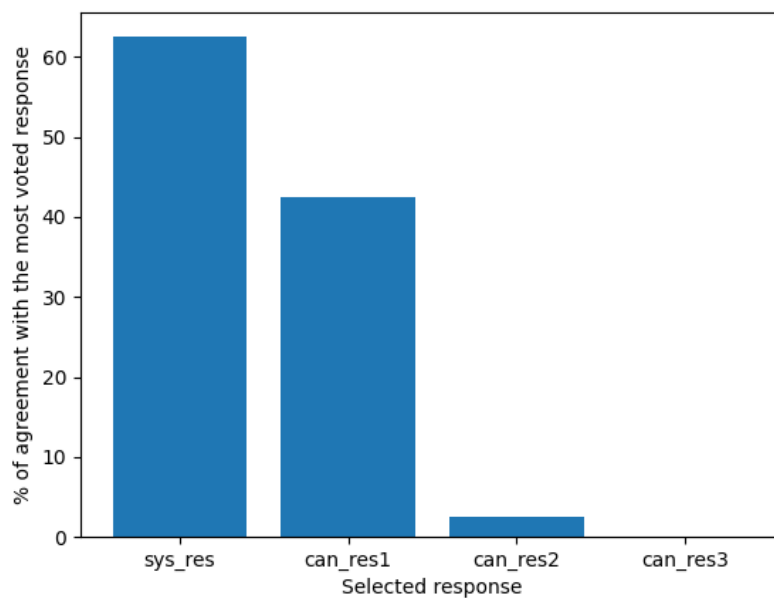


FIGURA 5.5: Porcentaje de coincidencia entre las respuestas candidatas y la respuesta más votada.

Capítulo 6

Conclusiones

En este capítulo se presentan las conclusiones del trabajo. También se delimitan las limitaciones del mismo. Finalmente, se describen las aportaciones de la investigación y se proponen algunas líneas de trabajo futuro.

6.1. Conclusiones

Actualmente existe un gran interés en desarrollar propuestas centradas en proporcionar terapia cognitiva y apoyo asistencial a los pacientes con demencia. Sin embargo, son pocos los estudios que se centran en el desarrollo de sistemas que proporcionen terapia de reminiscencia y los que lo hacen están desarrollados en su mayoría para el idioma inglés. En este trabajo se propone un sistema conversacional que puede ser empleado como herramienta para brindar terapia de reminiscencia a pacientes con demencia de tipo Alzheimer en etapa inicial en el idioma español.

La principal diferencia del sistema propuesto respecto a los trabajos relacionados en el estado del arte, radica en la capacidad del sistema para generar conversaciones personalizadas de acuerdo a la historia de vida, las preferencias y los intereses de cada paciente. Este enfoque brindaría a los usuarios del sistema los beneficios de la terapia de reminiscencia cuando se aplica de forma constante. Para que el sistema tenga la capacidad de conversar sobre distintos temas, se creó un conjunto de contextos de diálogos, los cuales brindan al sistema la posibilidad de brindar

terapia de reminiscencia sobre diferentes temas como: familia, lugares visitados, festividades favoritas, por mencionar algunos.

La evaluación realizada al sistema conversacional, nos permite afirmar que las respuestas brindadas por el sistema en los diferentes contextos de diálogo son adecuadas y por lo tanto podría conducir la terapia de reminiscencia con pacientes reales. Sin embargo, es importante mencionar que de acuerdo a las características de la demencia de tipo Alzheimer, los beneficios de brindar la terapia de reminiscencia no pueden ser generalizados para todos las personas que la padecen.

6.2. Limitaciones

Dentro del trabajo desarrollado existen ciertas limitaciones tanto del aspecto técnico como metodológico las cuales se mencionan a continuación.

En primer lugar, debido a que el sistema desarrollado emplea servicios de terceros para las tareas de reconocimiento automático del habla y de sintetizar el texto a habla (Google Cloud Speech-to-Text y Google Cloud Text-to-Speech), existe un límite en el número de transcripciones y emisiones que se pueden realizar de forma gratuita. Por lo que si se desea utilizar el sistema conversacional de forma recurrente con un considerable número de pacientes, esto representaría un costo en el uso de estos servicios.

Además, aunque los servicios implementados para estas tareas son soluciones con un buen nivel de desarrollo, existe una tasa de error al momento de desempeñar su función. Por lo que esto impacta de forma directa en el desempeño del sistema conversacional durante la interacción.

Otra limitación es que el sistema sólo tiene la capacidad de personalizar la conversación basándose en la información recolectada de los pacientes. Por lo que no tiene la capacidad de realiza ninguna búsqueda en fuentes externas para obtener información relacionada con el diálogo y que se pueda utilizar para enriquecer la conversación. Además, el sistema conversacional sólo puede realizar una sesión de terapia mediante voz. Sin embargo, en las sesiones realizadas por terapeutas humanos, comúnmente suelen emplearse objetos que permiten reforzar la terapia, como fotografías, música, juguetes u olores.

Respecto a la evaluación del sistema, una limitante relevante fue el número de evaluadores que participaron. Y si bien todos ellos tienen conocimiento y experiencia sobre la terapia de reminiscencia, resulta relevante incrementar el número de participante. Además, en esta etapa del trabajo se evaluó si los diálogos generados por el sistema son adecuados para llevar a cabo la terapia. Sin embargo, también es necesario evaluar el desempeño del sistema brindando la terapia en pacientes reales.

6.3. Aportaciones

Las principales aportaciones del presente trabajo de investigación son las siguientes:

- La creación de un conjunto de contextos de diálogos diseñados para ser utilizados en terapia de reminiscencia.
- Un método para personalización de conversaciones con PcA de acuerdo su historia de vida y preferencias.
- Un prototipo de sistema conversacional para brindar terapia de reminiscencia a personas con Alzheimer en etapa inicial.

A continuación se listan las publicaciones derivadas de este trabajo de investigación:

1. Maria J. Somodevilla, Victor Manuel Morales-de-Jesús. (2019). Towards a Natural Language-based Dialog Model as an Assistant Tool for Dementia Therapy. *Research in Computing Science*, 148(3): 195 – 203
2. Victor Manuel Morales-de-Jesús, Maria J. Somodevilla. (2020). A Conversational Model for the Reminiscence Therapy of Patients with Early Stage of Alzheimer. *Research in Computing Science*, 149(4): 57 – 67
3. Morales-de-Jesús, V.; Gómez-Adorno, H.; Somodevilla-García, M.; Vilariño, D. (2021) Conversational System as Assistant Tool in Reminiscence Therapy for People with Early-Stage of Alzheimer's. *Healthcare*, 1, 0.

6.4. Trabajo futuro

Como continuación de este trabajo de investigación de establecieron los siguientes puntos:

- Evaluación del sistema en pacientes reales. Si bien hasta esta etapa de la investigación, la evaluación del sistema para llevar a cabo la terapia ha sido realizada por personal relacionado al cuidado de PcA, es de relevante importancia que el sistema pueda ser utilizado por pacientes reales y realizar una evaluación en escenarios de conversación reales.
- Agregar el uso de fotografías y música en el sistema conversacional. Esto permitiría brindar una experiencia más enriquecedora a la PcA.
- Implementar funcionalidades adicionales al avatar como: movimiento, manifestación de emociones, entre otras.

Bibliografía

- Ali, M. R., Razavi, Z., Mamun, A. A., Langevin, R., Rawassizadeh, R., Schubert, L., y Hoque, M. E. (2018). A Virtual Conversational Agent for Teens with Autism: Experimental Results and Design Lessons. *arXiv preprint arXiv:1811.03046*.
- Alonso, F., Gorostiza, J., y Salichs, M. (2013). Preliminary experiments on hri for improvement the robotic dialog system (rds). En *Robocity2030 11th Workshop on Social Robots*.
- Alonso-Martín, F., Malfaz, M., Sequeira, J., Gorostiza, J. F., y Salichs, M. A. (2013). A multimodal emotion detection system during human–robot interaction. *Sensors*, 13(11):15549–15581.
- Alzheimer’s Association (2020). Stages of alzheimer’s. <https://www.alz.org/alzheimers-dementia/stages?lang=en-US>. Alzheimer’s Association, fecha de acceso: 20–Junio–2020.
- Alzheimer’s Disease International (2019). World Alzheimer Report 2019: Attitudes to Dementia. Technical report, Alzheimer’s Disease International, London UK.
- Austin, J. L. (1968). *How to do things with words*. Oxford University Press.
- Bach, K. y Harnish, R. (1979). *Linguistic Communication and Speech Acts*. MIT Press.
- Banerjee, S. y Lavie, A. (2005). Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. En *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pp. 65–72.

- Barrera-López, F., López-Beltrán, E., Baldivieso-Hurtado, N., Maple-Alvarez, I., López-Moraila, M., y Murillo-Bonilla, L. (2018). Diagnóstico Actual de la Enfermedad de Alzheimer. *Revista de Medicina Clínica*, 2(2):57–73.
- Becerra, A., de la Rosa, J. I., y González, E. (2017). Speech recognition in a dialog system: from conventional to deep processing. *Multimedia Tools and Applications*.
- Burtsev, M., Seliverstov, A., Airapetyan, R., Arkhipov, M., Baymurzina, D., Bushkov, N., Gureenkova, O., Khakhulin, T., Kuratov, Y., y Kuznetsov, D. (2018). Deepavlov: Open-source library for dialogue systems. En *Proceedings of ACL 2018, System Demonstrations*, pp. 122–127.
- Casey, D., Felzmann, H., Pegman, G., Kouroupetroglou, C., Murphy, K., Koumpis, A., y Whelan, S. (2016). What people with dementia want: designing mario an acceptable robot companion. En *International Conference on Computers Helping People with Special Needs*, pp. 318–325. Springer.
- Chinaei, H., Currie, L. C., Danks, A., Lin, H., Mehta, T., y Rudzicz, F. (2017). Identifying and avoiding confusion in dialogue with people with alzheimer’s disease. *Computational Linguistics*, 43(2):377–406.
- Choi, E., He, H., Iyyer, M., Yatskar, M., Yih, W.-t., Choi, Y., Liang, P., y Zettlemoyer, L. (2018). Quac: Question answering in context. *arXiv preprint arXiv:1808.07036*.
- Clark, A. y Popescu-Belis, A. (2004). Multi-level dialogue act tags. En *Proceedings of SIGDIAL ’04 (5th SIGDIAL Workshop on Discourse and Dialog)*. Cambridge.
- Cockrell, J. R. y Folstein, M. F. (2002). Mini-mental state examination. *Principles and practice of geriatric psychiatry*, pp. 140–141.
- Cruz-Sandoval, D. y Favela, J. (2017). Semi-autonomous Conversational Robot to Deal with Problematic Behaviors from People with Dementia. En Ochoa, Sergio F. and Singh, P. y Bravo, J., editores, *Ubiquitous Computing and Ambient Intelligence*, pp. 677–688, Cham. Springer International Publishing.
- Cruz-Sandoval, D., Favela, J., y Sandoval, E. B. (2018). Strategies to Facilitate the Acceptance of a Social Robot by People with Dementia. En *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, HRI ’18*, pp. 95–96, New York, NY, USA. ACM.

- Dean Webster, J. (2003). The reminiscence circumplex and autobiographical memory functions. *Memory*, 11(2):203–215.
- del Valle-Agudo, D., Calle, J., y Martínez, P. (2006). Enfoque Metodológico para Incorporar Conocimiento de Dominio a Sistemas de Diálogo Intencionales. *Procesamiento del Lenguaje Natural*, 37.
- Deriu, J., Rodrigo, A., Otegi, A., Echegoyen, G., Rosset, S., Agirre, E., y Cieliebak, M. (2021). Survey on evaluation methods for dialogue systems. *Artificial Intelligence Review*, 54(1):755–810.
- DeVault, D., Leuski, A., y Sagae, K. (2011). Toward learning and evaluation of dialogue policies with text examples. En *Proceedings of the SIGDIAL 2011 Conference*, pp. 39–48.
- Devlin, J., Chang, M.-W., Lee, K., y Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- Fernández, P. H., Velarde, S. I., Hernández, M. F., y Murguía, V. (2014). *Dinámica demográfica 1990-2010 y proyecciones de población 2010-2030*. Consejo Nacional de Población (CONAPO).
- Fernández, R., Ginzburg, J., y Lappin, S. (2007). Classifying non-sentential utterances in dialogue: A machine learning approach. *Computational Linguistics*, 33(3):397–427.
- Fitzpatrick, K. K., Darcy, A., y Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR mental health*, 4(2):e7785.
- Galley, M., Brockett, C., Sordani, A., Ji, Y., Auli, M., Quirk, C., Mitchell, M., Gao, J., y Dolan, B. (2015). deltableu: A discriminative metric for generation tasks with intrinsically diverse targets. *arXiv preprint arXiv:1506.06863*.
- Gandhe, S. y Traum, D. (2016). A semi-automated evaluation metric for dialogue model coherence. En *Situated Dialog in Speech-Based Human-Computer Interaction*, pp. 217–225. Springer.

- Gitlin, L. N., Hodgson, N., Jutkowitz, E., y Pizzi, L. (2010). The Cost-Effectiveness of a Nonpharmacologic Intervention for Individuals With Dementia and Family Caregivers: The Tailored Activity Program. *The American Journal of Geriatric Psychiatry*, 18(6):510–519.
- Gonzalez, J., Mayordomo, T., Torres, M., Sales, A., y Meléndez, J. C. (2015). Reminiscence and dementia: a therapeutic intervention. *International Psychogeriatrics*, 27(10):1731–1737.
- González-Arévalo, K. A. (2015). Terapia de reminiscencia y sus efectos en los pacientes mayores con demencia. *Psicogeriatría*, 5(3):101–111.
- Grosz, B. J. y Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3):175–204.
- Gunkel, D. J. (2016). Computational interpersonal communication: Communication studies and spoken dialogue systems. *communication+ 1*, 5(1):1–20.
- Gutierrez, L. (2014). Demencias en México. En *Demencias, una visión panorámica*, pp. 9–20. Universitaria Potosina.
- Hernández-Domínguez, L., Ratte, S., Davis, B., y Pope, C. (2016). Conversing with the elderly in latin america: a new cohort for multimodal, multilingual longitudinal studies on aging. En *Proceedings of the 7th Workshop on Cognitive Aspects of Computational Language Learning*, pp. 16–21.
- Heudin, J.-C. (2018). An emotional multi-personality architecture for intelligent conversational agents. En *Transactions on Computational Collective Intelligence XXVIII*, pp. 143–163. Springer.
- Inkster, B., Sarda, S., y Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: real-world data evaluation mixed-methods study. *JMIR Mhealth Uhealth*, 6(11):e12106.
- Irazoki, E., García-Casal, J. A., Sánchez-Meca, J., y Franco-Martín, M. (2017). Eficacia de la terapia de reminiscencia grupal en personas con demencia. revisión sistemática y metaanálisis. *Rev Neurol*, 65(10):447–456.
- Jadhav, K. P. y Thorat, S. A. (2020). Towards designing conversational agent systems. En *Computing in Engineering and Technology*, pp. 533–542. Springer.

- Ji, G. y Bilmes, J. (2006). Backoff model training using partially observed data: Application to dialog act tagging. En *Proceedings of the main conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics*, pp. 280–287. Association for Computational Linguistics.
- Kamita, T., Ito, T., Matsumoto, A., Munakata, T., e Inoue, T. (2019). A chatbot system for mental healthcare based on sat counseling method. *Mobile Information Systems*, 2019.
- Larsson, S., Bohlin, P., Bos, J., y Traum, D. (1999). Trindikit manual. Technical report, Tech. rept. Deliverable.
- Lawler, J. y Dry, H. A. (2008). *Using computers in linguistics: a practical guide*. Psychology Press.
- Lemon, O. (2011). Learning what to say and how to say it: Joint optimisation of spoken dialogue management and natural language generation. *Computer Speech & Language*, 25(2):210 – 221.
- Liddy, E. D. (1998). Enhanced text retrieval using natural language processing. *Bulletin of the Association for Information Science and Technology*, 24(4):14–16.
- Lin, C.-Y. (2004). Rouge: A package for automatic evaluation of summaries. En *Text summarization branches out*, pp. 74–81.
- Lippert, A., Shubeck, K., Morgan, B., Hampton, A., y Graesser, A. (2020). Multiple agent designs in conversational intelligent tutoring systems. *Technology, Knowledge and Learning*, 25(3):443–463.
- Litman, D., Hirschberg, J., y Swerts, M. (2006). Characterizing and predicting corrections in spoken dialogue systems. *Computational linguistics*, 32(3):417–438.
- Liu, C.-W., Lowe, R., Serban, I. V., Noseworthy, M., Charlin, L., y Pineau, J. (2016). How not to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation. *arXiv preprint arXiv:1603.08023*.
- Lizama, R. A. D. (2018). Género y Vejez en México, retos y oportunidades. *Pluralidad y Consenso*, 7(34):176–179.

- Llisterri, J., Carbó, C., Machuca, M., De la Mota, C., Riera, M., y Ríos, A. (2003). El papel de la fonética en el desarrollo de las tecnologías del habla. *Memorias de las VII Jornadas de Lingüística*. Cádiz, Servicio de Publicaciones de la Universidad de Cádiz.
- Mairittha, T., Mairittha, N., e Inoue, S. (2019). Evaluating a spoken dialogue system for recording systems of nursing care. *Sensors*, 19(17):3736.
- Marineau, J., Wiemer-Hastings, P., Harter, D., Olde, B., Chipman, P., Karnavat, A., Pomeroy, V., Rajan, S., y Graesser (2000). Classification of speech acts in tutorial dialog. En *Proceedings of the Workshop on Modeling Human Teaching Tactics and Strategies of ITS 2000*, pp. 65–71.
- Martinez-Hinarejos, C. D., Sanchis, E., Garcia-Granada, F., y Aibar, P. (2002). A labelling proposal to annotate dialogues. En *Proceedings of the 3rd LREC*, volumen 5, pp. 1566–1582. Citeseer.
- Mavridis, N. (2015). A review of verbal and non-verbal human–robot interactive communication. *Robotics and Autonomous Systems*, 63:22–35.
- Miao, Y., Gowayyed, M., y Metze, F. (2015). EESSEN: End-to-end speech recognition using deep RNN models and WFST-based decoding. En *2015 IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU)*, pp. 167–174. IEEE.
- Miller, G. A. (1995). Wordnet: a lexical database for English. *Communications of the ACM*, 38(11):39–41.
- Navarro, R. F., Rodríguez, M. D., y Favela, J. (2016). Use and Adoption of an Assisted Cognition System to Support Therapies for People with Dementia. *Computational and mathematical methods in medicine*, 2016.
- Olazarán, J., Reisberg, B., Clare, L., Cruz, I., Peña-Casanova, J., Del Ser, T., Woods, B., Beck, C., Auer, S., y Lai, C. (2010). Nonpharmacological therapies in alzheimer’s disease: a systematic review of efficacy. *Dementia and geriatric cognitive disorders*, 30(2):161–178.
- O’Shea, J., Bandar, Z., y Crockett, K. (2012). A multi-classifier approach to dialogue act classification using function words. En *Transactions on Computational Collective Intelligence VII*, pp. 119–143. Springer.

- Otter, D. W., Medina, J. R., y Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. *IEEE Transactions on Neural Networks and Learning Systems*.
- Pan, B., Yang, Y., Zhao, Z., Zhuang, Y., Cai, D., y He, X. (2019). Discourse marker augmented network with reinforcement learning for natural language inference. *arXiv preprint arXiv:1907.09692*.
- Papineni, K., Roukos, S., Ward, T., y Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. En *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*, pp. 311–318.
- Park, Y., Patwardhan, S., Visweswariah, K., y Gates, S. C. (2008). “An empirical analysis of word error rate and keyword error rate”. En *Ninth Annual Conference of the International Speech Communication Association*, pp. 2070–2073.
- Perera, R. y Nand, P. (2017). Recent advances in natural language generation: A survey and classification of the empirical literature. *Computing and Informatics*, 36(1):1–32.
- Phair, L. (2010). Caring for people with dementia in acute care settings. *Nursing Older People*, 22(1):9–10.
- Ramírez Sánchez, J. M., Montalvo Bereau, A. R., y Calvo de Lara, J. R. (2019). “Evaluación de Rasgos Acústicos para el Reconocimiento Automático del Habla en Escenarios Ruidosos usando Kaldi”. *Ingeniería Electrónica, Automática y Comunicaciones*, vol. 40:pp. 51 – 71.
- Reilly, R., Ferrari, G., y Prodanof, I. (1988). Framework for a model of dialogue. En *Proceedings of the 12th conference on Computational linguistics-Volume 2*, pp. 540–543. Association for Computational Linguistics.
- Rieser, V. y Lemon, O. (2009). Does this list contain what you were searching for? learning adaptive dialogue strategies for interactive question answering. *Natural Language Engineering*, 15(1):55–72.
- Rokkaku, R. (2005). Care management sheet pack for the elderly with dementia: The center method ver. 03. *Nihon Ronen Igakkai zasshi. Japanese journal of geriatrics*, 42(3):318–319.

- Rudzicz, F., Wang, R., Begum, M., y Mihailidis, A. (2015). Speech Interaction with Personal Assistive Robots Supporting Aging at Home for Individuals with Alzheimer's Disease. *ACM Transactions on Accessible Computing*, 7(2):1–22.
- Russo, A., D'Onofrio, G., Gangemi, A., Giuliani, F., Mongiovi, M., Ricciardi, F., Greco, F., Cavallo, F., Dario, P., y Sancarolo, D. (2019). Dialogue systems and conversational agents for patients with dementia: The human–robot interaction. *Rejuvenation research*, 22(2):109–120.
- Sakakibara, S., Saiki, S., Nakamura, M., y Yasuda, K. (2017). Generating personalized dialogue towards daily counseling system for home dementia care. En Duffy, V. G., editor, *Digital Human Modeling. Applications in Health, Safety, Ergonomics, and Risk Management: Health and Safety*, pp. 161–172, Cham. Springer International Publishing.
- Sarabia, M., Young, N., Canavan, K., Edginton, T., Demiris, Y., y Vizcaychipi, M. P. (2018). Assistive Robotic Technology to Combat Social Isolation in Acute Hospital Settings. *International Journal of Social Robotics*.
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*, volumen 626. Cambridge University Press.
- Searle, J. R. (1975). Indirect speech acts. *Syntaxis and Semantics*, 3:59–82.
- Singh, K., Lytra, I., Radhakrishna, A. S., Shekarpour, S., Vidal, M.-E., y Lehmann, J. (2020). No one is perfect: Analysing the performance of question answering components over the dbpedia knowledge graph. *Journal of Web Semantics*, 65:100594.
- Small, J. A. y Gutman, G. (2002). Recommended and reported use of communication strategies in alzheimer caregiving. *Alzheimer Disease & Associated Disorders*, 16(4):270–278.
- Sosnowski, T. y Yordanova, K. (2020). A probabilistic conversational agent for intelligent tutoring systems. En *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, pp. 1–7.
- Stent, A. (2000). Rhetorical structure in dialog. En *Proceedings of the first international conference on Natural language generation*, volumen 14, pp. 247–252. Association for Computational Linguistics.

- Tanaka, H., Adachi, H., Ukita, N., Ikeda, M., Kazui, H., Kudo, T., y Nakamura, S. (2017). Detecting dementia through interactive computer avatars. *IEEE journal of translational engineering in health and medicine*, 5:1–11.
- Uchiya, T., Nishimura, R., Hirano, T., y Sakurai, M. (2018). Design of Reminiscence Therapy System for Elderly People with Dementia. En Barolli, L., Xha-fa, F., y Conesa, J., editores, *Advances on Broad-Band Wireless Computing, Communication and Applications*, pp. 844–853, Cham. Springer International Publishing.
- Voorhees, E. M. (2003). Overview of the trec 2003 robust retrieval track. En *Trec*, pp. 69–77.
- Voorhees, E. M. (2008). *Evaluating Question Answering System Performance*, pp. 409–430. Springer Netherlands, Dordrecht.
- Wallace, R. (2001). Artificial linguistic internet computer entity (ALICE).
- Weik, M. H. (2001). Natural Language Understanding. En *Computer Science and Communications Dictionary*, pp. 1073–1073. Springer US, Boston, MA.
- 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.
- Wolters, M. K., Kelly, F., y Kilgour, J. (2016). Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia. *Health Informatics Journal*, 22(4):854–866.
- Woods, B., O’Philbin, L., Farrell, E. M., Spector, A. E., y Orrell, M. (2018). Reminiscence therapy for dementia. *Cochrane database of systematic reviews*, (3).
- World Health Organization (2020). Dementia. <https://www.who.int/news-room/fact-sheets/detail/dementia>. World Health Organization, fecha de acceso: 20–Noviembre–2020.
- Wu, B., Li, K., Ge, F., Huang, Z., Yang, M., Siniscalchi, S. M., y Lee, C.-H. (2017). An end-to-end deep learning approach to simultaneous speech dereverberation and acoustic modeling for robust speech recognition. *IEEE Journal of Selected Topics in Signal Processing*, 11(8):1289–1300.

- Xu, J. y Lee, T. C. (2017). System and method for distributed text-to-speech synthesis and intelligibility. US Patent 9,761,219.
- Yampolskiy, R. V. (2013). Turing test as a defining feature of AI-completeness. En *Artificial intelligence, evolutionary computing and metaheuristics*, pp. 3–17. Springer.
- Yáñez, A. (2013). El enunciado y el contexto enunciativo: hacia la pragmática. *Revista Comunicación*, 11(2).
- Yasuda, K., Kuwahara, N., Kuwabara, K., Morimoto, K., y Tetsutani, N. (2013). Daily assistance for individuals with dementia via videophone. *American Journal of Alzheimer's Disease and other Dementias*, 28(5):508–516.
- Young, S., Gašić, M., Keizer, S., Mairesse, F., Schatzmann, J., Thomson, B., y Yu, K. (2010). The Hidden Information State model: A practical framework for POMDP-based spoken dialogue management. *Computer Speech & Language*, 24(2):150–174.
- Zhou, L., Gao, J., Li, D., y Shum, H.-Y. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, 46(1):53–93.
- Zubizarreta-Aizpuru, J. R. (1993). Un modelo de diálogo para diálogos orientados por la tarea. *Procesamiento del lenguaje natural*. N. 13; pp. 289-301.