



# **Aplicación de Reinforcement Learning y Transformers para el Aprendizaje de Idiomas**

Trabajo Fin de Máster,  
Convocatoria Marzo 2025

**Máster Propio en Inteligencia Artificial**

Segunda Edición,  
Curso Académico 2024

Por el alumno/a

**Julio Emanuel Suriano Bryk**

Con DNI 37.620.411

Dirigido por

**José Gabriel García Pardo**



*Education is not about thinning the herd.  
Education is about helping every student succeed.*

Andrew Ng



# Agradecimientos

Me gustaría agradecer a los profesores y compañeros del máster por su apoyo, sus valiosas aportaciones y los intercambios de ideas que han enriquecido este trabajo.

También quiero destacar a mi pareja por su comprensión y apoyo durante todo el proceso, lo que me permitió mantenerme enfocado y motivado.

Por último, agradezco a mi familia y amigos por su apoyo constante y por estar siempre disponibles para ofrecerme su ayuda y perspectiva.

# Índice general

Índice de figuras . . . . .	III
Índice de tablas . . . . .	IV
Índice de algoritmos . . . . .	V
Resumen . . . . .	1
1 Introducción . . . . .	2
1.1 Motivación . . . . .	2
1.1.1 Limitaciones Actuales . . . . .	2
1.1.2 Oportunidades de Mejora . . . . .	3
1.2 Objetivos . . . . .	3
1.2.1 Objetivo General . . . . .	3
1.2.2 Objetivo Específicos . . . . .	3
2 Estado del arte . . . . .	5
2.1 Sistemas de Aprendizaje Adaptativo . . . . .	5
2.2 Aplicaciones de LLM . . . . .	6
2.2.1 Asistentes y Diálogo . . . . .	6
2.2.2 Análisis y Corrección . . . . .	6
2.3 Tecnologías Emergentes con Compañeros de Aprendizaje . . . . .	6
2.4 Avances en Procesamiento de Voz . . . . .	7
2.5 Agentic AI . . . . .	7
2.6 Frameworks de Aprendizaje por Refuerzo . . . . .	8
2.7 Aplicación de las Tecnologías en el Trabajo . . . . .	9
3 Marco teórico . . . . .	10
3.1 Fundamentos del Aprendizaje de Idiomas . . . . .	10
3.1.1 Teorías de Adquisición del Lenguaje . . . . .	10
3.1.2 Factores que Influyen en el Aprendizaje de Segundas Lenguas . . . . .	11
3.1.3 Metodologías de Enseñanza . . . . .	11



3.1.4	Métodos Tradicionales . . . . .	11
3.1.5	Enfoques Modernos . . . . .	12
3.1.6	Desafíos en la Personalización del Aprendizaje . . . . .	12
3.1.7	Evaluación del Progreso . . . . .	13
3.2	Inteligencia Artificial en Educación . . . . .	13
3.2.1	Evolución de los Sistemas de Aprendizaje Adaptativo . . . . .	13
3.2.2	Arquitecturas de Sistemas Educativos Inteligentes . . . . .	14
3.2.3	Personalización y Adaptación Dinámica . . . . .	14
3.2.4	Métodos de Evaluación Automática . . . . .	14
3.2.5	Sistemas de Recomendación Educativa . . . . .	15
3.3	Procesamiento del Lenguaje Natural y LLMs . . . . .	15
3.3.1	Arquitectura Transformer . . . . .	16
3.3.2	Large Language Models (LLMs) . . . . .	16
3.3.3	Sistemas de Recuperación Aumentada con Generación (RAG) . . . . .	17
3.3.4	Aplicaciones y Ventajas de RAG en Educación . . . . .	18
3.4	Aprendizaje por Refuerzo . . . . .	18
3.4.1	Fundamentos Teóricos del RL . . . . .	18
3.4.2	Proximal Policy Optimization (PPO) . . . . .	18
3.4.3	Evaluación de Políticas de Aprendizaje . . . . .	20
3.5	Tecnologías de Procesamiento de Voz . . . . .	21
3.5.1	Reconocimiento Automático del Habla (STT) . . . . .	21
3.5.2	Síntesis de Voz (TTS) . . . . .	22
3.5.3	Integración en Sistemas de Aprendizaje . . . . .	22
4	Material . . . . .	24
5	Métodos . . . . .	25
6	Resultados . . . . .	26
7	Conclusiones . . . . .	27
	Glosario . . . . .	29
	Glosario . . . . .	31
A	Apéndice A . . . . .	32
	Bibliografía . . . . .	33

# Índice de figuras

3.1	Flujo de información en un sistema RAG . . . . .	17
-----	--	----





# Índice de tablas

# Índice de algoritmos

1	Algoritmo <i>Proximal Policy Optimization</i> (PPO) . . . . .	19
---	---	----

# Resumen

El sistema propuesto en este Trabajo de Fin de Máster integra [Reinforcement Learning \(RL\)](#), arquitecturas [Transformers](#) y un sistema de [Retrieval-Augmented Generation \(RAG\)](#) para optimizar el aprendizaje de idiomas. El sistema recomienda ejercicios personalizados y genera diálogos interactivos con retroalimentación en tiempo real. Además, genera rutas de aprendizaje adaptativas y simulaciones de conversación, integrando tecnologías de [Text-to-Speech \(TTS\)](#) y [Speech-to-Text \(STT\)](#) para el desarrollo de habilidades auditivas y de pronunciación. El sistema evoluciona continuamente mediante el análisis de datos de interacción, registros de audio y diálogos simulados, adaptándose a las necesidades específicas de cada estudiante.

# Introducción

# 1

El aprendizaje de idiomas en la era digital ha experimentado una transformación significativa gracias a los avances en el sector de [Inteligencia Artificial \(IA\)](#). Sin embargo, uno de los mayores desafíos sigue siendo la personalización efectiva del proceso de aprendizaje para adaptarse a las necesidades individuales de cada estudiante. Este trabajo propone un enfoque innovador que combina técnicas de [RL](#) con arquitecturas [Transformers](#) introducidas y tecnologías de procesamiento de voz para crear un sistema de aprendizaje de idiomas adaptativo y personalizado.

## 1.1. Motivación

La adquisición de una segunda lengua es un proceso complejo que varía significativamente entre individuos. Los métodos tradicionales de enseñanza de idiomas, incluso en su forma digitalizada, presentan limitaciones significativas que impiden una personalización efectiva y una adaptación dinámica al progreso del estudiante.

### 1.1.1. Limitaciones Actuales

En la actualidad, los métodos de enseñanza de idiomas enfrentan varias limitaciones que afectan la eficacia del aprendizaje. Estas limitaciones pueden clasificarse en cuatro categorías principales:

- **Rigidez Estructural:** Los programas siguen secuencias predefinidas que no se adaptan al progreso real del estudiante, limitando la capacidad de responder a sus necesidades específicas.
- **Falta de Personalización:** No consideran adecuadamente los diferentes estilos de aprendizaje, intereses y preferencias individuales, lo que puede afectar la motivación y la eficacia del aprendizaje.
- **Retroalimentación Limitada:** La mayoría de los sistemas proporcionan feedback básico sin considerar el contexto completo del aprendizaje, lo que dificulta la identificación de áreas de mejora específicas.

- **Práctica Conversacional Artificial:** Las interacciones suelen ser mecánicas y no reflejan la naturaleza dinámica del lenguaje real, lo que limita la capacidad del estudiante para aplicar sus habilidades en situaciones de la vida real.

Estas limitaciones resaltan la necesidad de un enfoque más flexible y personalizado en la enseñanza de idiomas, que pueda adaptarse a las necesidades y progresos individuales de cada estudiante, proporcionando una experiencia de aprendizaje más efectiva y motivadora.

### 1.1.2. Oportunidades de Mejora

A pesar de los avances en la enseñanza de idiomas, existen varias áreas donde se pueden realizar mejoras significativas:

- **Adaptabilidad Dinámica:** Implementar sistemas que ajusten el contenido y la dificultad en tiempo real, basándose en el rendimiento y las necesidades del estudiante.
- **Personalización Profunda:** Considerar múltiples factores individuales, como el estilo de aprendizaje, intereses y ritmo de progreso, para optimizar el proceso de aprendizaje.
- **Interacción Natural:** Utilizar tecnologías avanzadas, como modelos de lenguaje natural y procesamiento de voz, para simular conversaciones más realistas y dinámicas.
- **Feedback Contextual:** Proporcionar retroalimentación detallada y específica, basada en el contexto y el perfil del estudiante, para mejorar la comprensión y el rendimiento.

## 1.2. Objetivos

### 1.2.1. Objetivo General

Desarrollar un sistema de aprendizaje de idiomas que utilice [RL](#), [Transformers](#) y un [Sistema Multi-Agente](#) para una experiencia de aprendizaje personalizada, adaptativa y efectiva.

### 1.2.2. Objetivo Específicos

Para alcanzar el objetivo general, se han definido varios objetivos específicos que se centran en la implementación de técnicas avanzadas de inteligencia artificial. Estos objetivos específicos se enfocan en la aplicación de [RL](#) para optimizar el proceso de aprendizaje, el uso de modelos [Large Language Model \(LLM\)](#) para mejorar la interacción, el desarrollo de herramientas basadas en [TTS](#) y [STT](#) para perfeccionar las habilidades lingüísticas, y la integración de sistemas [RAG](#) para una gestión eficiente del conocimiento. A continuación, se detallan los objetivos específicos:

### 1.2.2.1. Optimización del Aprendizaje

El objetivo de esta sección es implementar estrategias que optimicen el proceso de aprendizaje mediante la personalización y la evaluación continua. Para ello, se propone implementar un sistema de [RL](#) que optimice rutas de aprendizaje personalizadas, desarrollar mecanismos de adaptación dinámica del contenido y crear sistemas de evaluación continua del progreso.

### 1.2.2.2. Mejora de la Interacción

El objetivo de esta sección es mejorar la interacción entre el sistema y el usuario mediante el uso de [LLM](#). Para ello, se propone integrar modelos [LLM](#) para el [Natural Language Processing \(NLP\)](#), desarrollar sistemas de diálogo contextuales e implementar análisis de errores en tiempo real. Estas mejoras permitirán una comunicación más fluida y natural, facilitando una experiencia de aprendizaje más efectiva y personalizada.

### 1.2.2.3. Perfeccionamiento de Habilidades Lingüísticas

El objetivo aquí es desarrollar herramientas que ayuden a los usuarios a mejorar sus habilidades lingüísticas, especialmente en pronunciación y comprensión. Para lograr esto, se propone crear sistemas de evaluación de pronunciación usando [TTS](#) y [STT](#), desarrollar ejercicios adaptativos de comprensión e implementar práctica conversacional contextual.

### 1.2.2.4. Gestión del Conocimiento

Esta sección se dedica a la integración y gestión de recursos educativos para proporcionar un acceso eficiente y actualizado a la información. Se busca integrar sistemas [RAG](#) para el acceso a recursos educativos, desarrollar bases de conocimiento dinámicas e implementar mecanismos de actualización de contenido.

# Estado del arte

# 2

Esta sección presenta una revisión detallada de las tecnologías y sistemas más avanzados en el campo del aprendizaje de idiomas asistido por IA. Se exploran los sistemas de aprendizaje adaptativo, destacando innovaciones recientes en plataformas populares. Además, se analizan las aplicaciones LLM en la generación de diálogos y corrección de textos. También se presentan tecnologías emergentes con compañeros de aprendizaje basados en IA, avances en procesamiento de voz (TTS y STT), y la implementación de tecnologías de Sistema Multi-Agente. Finalmente, se revisan los frameworks de RL más utilizados en la industria.

## 2.1. Sistemas de Aprendizaje Adaptativo

Los sistemas modernos de aprendizaje de idiomas han evolucionado significativamente con la integración de IA y aprendizaje automático. A continuación, se destacan algunas innovaciones recientes:

- **Busuu Conversations (2024)**<sup>1</sup>: Incorpora un sistema de IA que analiza patrones de error y ajusta dinámicamente el contenido para mejorar la eficacia del aprendizaje.
- **Duolingo Max (2024)**<sup>2</sup>: Utiliza GPT-4 para generar explicaciones personalizadas y mantener conversaciones contextuales, adaptándose al nivel del usuario.
- **Babbel Everyday Conversations (2023)**<sup>3</sup>: Combina IA con tutores humanos para optimizar la experiencia de aprendizaje híbrido, ofreciendo una interacción más personalizada.
- **Lingvist (2023)**<sup>4</sup>: utiliza datos contextuales para generar ejercicios, lecciones y recomendaciones adaptadas, facilitando la recuperación de contenidos lingüísticos relevantes y la generación de actividades interactivas.

---

<sup>1</sup><https://www.busuu.com>

<sup>2</sup><https://www.duolingo.com>

<sup>3</sup><https://www.babbel.com>

<sup>4</sup><https://www.lingvist.com>

## 2.2. Aplicaciones de LLM

Los LLM han revolucionado el aprendizaje de idiomas en múltiples aspectos, proporcionando herramientas avanzadas para la generación de diálogos y el análisis y corrección de textos.

### 2.2.1. Asistentes y Diálogo

- **ChatGPT (2022)**<sup>5</sup>: Revolucionó la interacción humano-IA estableciendo el estándar de interfaces conversacionales naturales y creando un ecosistema completo de desarrollo.
- **Claude (2023)**<sup>6</sup>: Destacó por su precisión superior en análisis de documentos y capacidad de seguir instrucciones complejas con menor tendencia a la alucinación.
- **Azure Language Studio (2023)**<sup>7</sup>: Ofrece herramientas de análisis lingüístico y generación de contenido educativo, mejorando la calidad del aprendizaje.
- **LLaMA (2023)**<sup>8</sup>: Modelo de Código Abierto desarrollado por Meta, diseñado para ser eficiente y accesible para la investigación y aplicaciones prácticas.

### 2.2.2. Análisis y Corrección

- **Grammarly with GrammarlyGO (2023)**<sup>9</sup>: Utiliza IA generativa para proporcionar correcciones contextuales y sugerencias de mejora, ayudando a los usuarios a escribir con mayor precisión.
- **DeepL Write (2023)**<sup>10</sup>: Sistema de corrección que considera el contexto cultural y el registro lingüístico, ofreciendo sugerencias más relevantes y precisas.

## 2.3. Tecnologías Emergentes con Compañeros de Aprendizaje

Los compañeros de aprendizaje basados en IA están emergiendo como una herramienta fundamental en la educación moderna. Estos asistentes virtuales proporcionan apoyo personalizado y adaptativo, actuando como tutores disponibles 24/7. A continuación, se presentan algunas implementaciones destacadas:

- **Khanmigo (2024)**<sup>11</sup>: Tutor virtual de Khan Academy que actúa como compañero de estudio personalizado, proporcionando explicaciones adaptativas, guía paso a paso y retroalimentación instantánea en múltiples materias.

<sup>5</sup><https://chatgpt.com/>

<sup>6</sup><https://claude.ai/>

<sup>7</sup><https://language.cognitive.azure.com/>

<sup>8</sup><https://ai.facebook.com/blog/large-language-model-llama>

<sup>9</sup><https://www.grammarly.com>

<sup>10</sup><https://www.deepl.com/write>

<sup>11</sup><https://www.khanacademy.org/khan-labs>



- **Third Space Learning (2024)**<sup>12</sup>: Plataforma que combina tutores humanos con IA para crear una experiencia de aprendizaje híbrida, donde el sistema analiza las interacciones y proporciona insights personalizados.
- **Riiid SANTA (2023)**<sup>13</sup>: Sistema de tutoría adaptativa para predecir el rendimiento del estudiante y personalizar el contenido, maximizando la eficiencia del aprendizaje mediante análisis predictivo.

### 2.4. Avances en Procesamiento de Voz

Las tecnologías de TTS y STT han mejorado en naturalidad y expresividad, proporcionando voces más humanas y adaptativas, así como una transcripción precisa y rápida. A continuación, se presentan algunas de las tecnologías más destacadas en este campo:

- **Whisper OpenAI (2022)**<sup>14</sup>: Reconocimiento de voz multilingüe de alta precisión, eficaz en ambientes ruidosos y con diversos acentos. Es Código Abierto y se utiliza para transcripción automática y análisis de voz en múltiples idiomas.
- **Google Speech-to-Text/Text-to-Speech (2023)**<sup>15</sup>: Reconocimiento de voz en tiempo real con alta precisión, soporte para múltiples idiomas y fácil integración con otras plataformas de Google. Comúnmente usado en asistentes virtuales y transcripción de reuniones en vivo.
- **Microsoft Azure AI Speech (2023)**<sup>16</sup>: Transcripción precisa y rápida, con capacidades avanzadas de personalización y adaptación al contexto. Ideal para sistemas de atención al cliente y análisis de conversaciones en tiempo real.
- **Deepgram (2023)**<sup>17</sup>: Plataforma de reconocimiento de voz basada en redes neuronales profundas, conocida por su rapidez y precisión. Utilizada para transcripción de llamadas y análisis de conversaciones de negocio.

### 2.5. Agentic AI

La tecnología de Sistema Multi-Agente se está convirtiendo en un área clave de innovación en el aprendizaje de idiomas. Estas tecnologías permiten la creación de agentes autónomos que pueden interactuar entre sí y con los usuarios para proporcionar experiencias de aprendizaje más dinámicas y personalizadas.

---

<sup>12</sup><https://thirdspacelearning.com>

<sup>13</sup><https://riiid.com>

<sup>14</sup><https://openai.com/research/whisper>

<sup>15</sup><https://cloud.google.com/speech-to-text>

<sup>16</sup><https://azure.microsoft.com/en-us/products/ai-services/ai-speech>

<sup>17</sup><https://deepgram.com>

- **LangChain (2022)**<sup>18</sup>: Plataforma [Código Abierto](#) que facilita la creación de [Sistema Multi-Agente](#). LangChain permite la integración de diferentes modelos de lenguaje y agentes especializados para tareas específicas, mejorando la interacción y la adaptabilidad del sistema.
- **CrewAI (2023)**<sup>19</sup>: Sistema multi-agente [Código Abierto](#) diseñado para la colaboración en equipo, permitiendo a los usuarios trabajar juntos en proyectos de aprendizaje de idiomas y recibir retroalimentación en tiempo real.
- **phiData (2023)**<sup>20</sup>: Plataforma [Código Abierto](#) que utiliza agentes especializados para analizar datos lingüísticos y proporcionar recomendaciones personalizadas para mejorar el aprendizaje de idiomas.
- **Autogen de Microsoft (2023)**<sup>21</sup>: Tecnología [Código Abierto](#) de Microsoft que permite la creación de agentes autónomos para tareas específicas en el aprendizaje de idiomas, mejorando la personalización y la eficacia del proceso educativo.

## 2.6. Frameworks de Aprendizaje por Refuerzo

El [RL](#) ha ganado popularidad en la industria debido a su capacidad para resolver problemas complejos mediante la optimización de políticas a través de la interacción con el entorno. A continuación, se presentan algunos de los frameworks de [RL](#) más utilizados en la industria, todos ellos [Código Abierto](#):

- **TensorFlow Agents (2019)**<sup>22</sup>: Una biblioteca de [RL](#) basada en [TensorFlow](#) que proporciona herramientas para construir, entrenar y evaluar agentes de [RL](#). Es compatible con una amplia gama de algoritmos y entornos.
- **Stable Baselines3 (2020)**<sup>23</sup>: Una implementación de algoritmos de [RL](#) en [PyTorch](#), diseñada para ser fácil de usar y extender. Es ampliamente utilizada para experimentación y desarrollo de soluciones de [RL](#).
- **TorchRL (2022)**<sup>24</sup>: Un framework de aprendizaje por refuerzo basado en [PyTorch](#), diseñado para ser flexible y fácil de usar. Proporciona herramientas para construir, entrenar y evaluar agentes de [RL](#) en diversos entornos.

---

<sup>18</sup><https://www.langchain.com>

<sup>19</sup><https://www.crewai.com>

<sup>20</sup><https://www.phidata.com>

<sup>21</sup><https://www.microsoft.com/en-us/research/project/autogen>

<sup>22</sup><https://www.tensorflow.org/agents>

<sup>23</sup><https://stable-baselines3.readthedocs.io>

<sup>24</sup><https://github.com/pytorch/rl>

## 2.7. Aplicación de las Tecnologías en el Trabajo

En este trabajo, se toman en cuenta las siguientes tecnologías y teorías para desarrollar un sistema de aprendizaje de idiomas asistido por [IA](#):

- **Sistemas de Aprendizaje Adaptativo:** Se implementa un sistema que analiza los patrones de error de los usuarios y ajusta dinámicamente el contenido.
- **Aplicaciones de LLM:** Se utiliza un [LLM](#) para generar diálogos y proporcionar correcciones contextuales.
- **Tecnologías Emergentes con Compañeros de Aprendizaje:** Se desarrolla un asistente virtual que actúa como compañero de aprendizaje, proporcionando apoyo personalizado y adaptativo.
- **Avances en Procesamiento de Voz:** Se integra tecnología de [TTS](#) y [STT](#) para mejorar la interacción del usuario con el sistema.
- **Agentic AI:** Se explora la creación de agentes autónomos que interactúan entre sí y con los usuarios para proporcionar experiencias de aprendizaje más dinámicas.
- **Frameworks de Aprendizaje por Refuerzo:** Se utilizan frameworks de [RL](#) para optimizar el proceso de aprendizaje y adaptar el contenido a las necesidades de los usuarios.

Estas tecnologías y teorías son fundamentales para el desarrollo de un sistema de aprendizaje de idiomas que es adaptativo, interactivo y altamente personalizado, mejorando significativamente la experiencia del usuario.

# Marco teórico

# 3

Este capítulo proporciona una base teórica para el aprendizaje de idiomas asistido por IA. Se exploran las teorías fundamentales de adquisición del lenguaje, los factores que influyen en el aprendizaje de segundas lenguas, y las metodologías de enseñanza. Además, se discuten los desafíos en la personalización del aprendizaje y los métodos de evaluación del progreso. Finalmente, se analiza la integración de la IA en la educación, incluyendo sistemas de aprendizaje adaptativo, procesamiento del lenguaje natural y aprendizaje por refuerzo.

## 3.1. Fundamentos del Aprendizaje de Idiomas

El aprendizaje de idiomas es un proceso complejo que involucra una variedad de factores cognitivos, sociales y afectivos. Esta sección explora las teorías fundamentales de adquisición del lenguaje, los factores que influyen en el aprendizaje de segundas lenguas, y las metodologías de enseñanza más efectivas.

### 3.1.1. Teorías de Adquisición del Lenguaje

El campo de la [Second Language Acquisition \(SLA\)](#) ha evolucionado significativamente en las últimas décadas, pasando de enfoques conductistas a perspectivas más cognitivas y socioculturales, y más recientemente, hacia la integración de tecnologías de IA y sistemas adaptativos que prometen revolucionar la manera en que se aprenden los idiomas.

Entre las teorías más influyentes en la adquisición de segundos idiomas, destaca especialmente el trabajo de [Krashen \(1982\)](#), quien desarrolló el Modelo del Monitor. Este modelo incluye cinco hipótesis fundamentales, siendo la más relevante la hipótesis del input comprensible, que establece que la adquisición ocurre cuando los estudiantes reciben input ligeramente por encima de su nivel actual de competencia.

Por su parte, [Ellis \(1994\)](#) propone un marco teórico más integrador, enfatizando la interacción entre factores cognitivos y ambientales en el aprendizaje de idiomas. Su trabajo destaca la importancia de considerar tanto los procesos mentales internos como las variables contextuales que influyen en la adquisición del lenguaje, proporcionando una base teórica sólida para entender cómo los estudiantes procesan y adquieren una segunda lengua.

### 3.1.2. Factores que Influyen en el Aprendizaje de Segundas Lenguas

Ellis (1994) identifica diversos factores que afectan el aprendizaje de idioma, que se pueden clasificar en internos y externos.

Los factores internos incluyen la edad del aprendiz, la aptitud lingüística, la motivación y actitud, los estilos cognitivos y las estrategias de aprendizaje, así como los rasgos de personalidad. La edad influye en la plasticidad cerebral y la capacidad de adquisición natural del lenguaje, mientras que la aptitud lingüística varía entre individuos y puede predecir el éxito en el aprendizaje. La motivación puede ser intrínseca o extrínseca, y los estilos cognitivos y estrategias de aprendizaje determinan cómo se procesa y retiene la información. Los rasgos de personalidad, como la extroversión, afectan la disposición a participar en interacciones comunicativas.

Por otro lado, los factores externos incluyen el contexto social y cultural, la exposición al idioma objetivo y la calidad y cantidad de input. El entorno de aprendizaje y el contexto socio-cultural influyen significativamente en las actitudes hacia la lengua objetivo y sus hablantes, determinando en gran medida el éxito del aprendizaje.

La exposición frecuente y variada al idioma es fundamental para desarrollar la competencia lingüística, y el input debe ser comprensible pero desafiante, siguiendo el principio de  $i+1$  de Krashen (1982). Este principio sugiere que el aprendizaje óptimo ocurre cuando el estudiante se expone a contenido ligeramente por encima de su nivel actual de competencia.

Además, factores como el estatus socioeconómico, el acceso a recursos educativos y tecnológicos, y las políticas lingüísticas del entorno también influyen significativamente en el proceso de aprendizaje. La disponibilidad de materiales auténticos y herramientas tecnológicas modernas puede enriquecer considerablemente la experiencia de aprendizaje y facilitar la exposición al idioma objetivo en contextos significativos.

### 3.1.3. Metodologías de Enseñanza

La evolución de las metodologías de enseñanza refleja nuestra comprensión cambiante del proceso de aprendizaje de idiomas:

#### 3.1.4. Métodos Tradicionales

El Método Gramática-Traducción, predominante durante el siglo XIX y principios del XX (Richards y Rodgers (2000)), se centra en el análisis detallado de reglas gramaticales y la traducción de textos. Este método enfatiza la precisión gramatical y la comprensión lectora, aunque ha sido criticado por su limitada atención a las habilidades comunicativas orales.

El Método Directo, introducido por Gouin (1892), surgió como respuesta a las limitaciones del método anterior, promoviendo la inmersión total en la lengua objetivo y evitando el uso de la lengua materna. Este enfoque enfatiza la importancia de la comunicación oral y la asociación directa entre el lenguaje y el significado, sin recurrir a la traducción.

El Método Audiolingüal, desarrollado durante la Segunda Guerra Mundial y fundamentado por Fries (1945), se basa en principios conductistas y enfatiza la formación de hábitos

lingüísticos a través de la repetición y el refuerzo. Este método utiliza ejercicios de patrón y diálogos memorizados para desarrollar automatismos en el uso del lenguaje.

#### 3.1.5. Enfoques Modernos

El Enfoque Comunicativo de la Enseñanza de Lenguas [Hymes \(1972\)](#) marcó un cambio revolucionario en la enseñanza de idiomas al enfatizar la competencia comunicativa sobre la mera precisión gramatical. Este enfoque transformó fundamentalmente la manera en que se enseñan los idiomas, priorizando las interacciones significativas y el uso del lenguaje en contextos reales.

El Aprendizaje Basado en Tareas [Nunan \(1989\)](#) representa otro pilar fundamental, organizando el aprendizaje alrededor de actividades comunicativas auténticas. Su efectividad radica en promover el aprendizaje natural del lenguaje mientras los estudiantes se enfocan en completar tareas prácticas y significativas.

El Aprendizaje Integrado de Contenidos y Lenguas [Coyle et al. \(2010\)](#) ha demostrado ser particularmente efectivo al integrar el aprendizaje de contenido académico con la adquisición del idioma. Este enfoque dual no solo mejora la eficiencia del aprendizaje sino que también aumenta significativamente la motivación de los estudiantes al proporcionar un contexto relevante y propósito claro para el uso del idioma.

#### 3.1.6. Desafíos en la Personalización del Aprendizaje

La personalización del aprendizaje representa uno de los mayores retos en la enseñanza de idiomas. Como señala [Ellis \(1994\)](#), un primer desafío fundamental es la identificación precisa del nivel del estudiante, que requiere evaluaciones comprehensivas que consideren no solo el conocimiento gramatical y léxico, sino también las habilidades comunicativas en diversos contextos.

La adaptación del contenido a diferentes estilos de aprendizaje constituye otro reto significativo, pues implica desarrollar materiales y actividades que satisfagan las preferencias y necesidades individuales de los estudiantes, considerando sus diferentes formas de procesar y retener la información lingüística. [Krashen \(1982\)](#) enfatiza la importancia de proporcionar input comprensible adaptado al nivel individual de cada estudiante.

El mantenimiento de la motivación requiere un equilibrio delicado entre desafío y apoyo, necesitando estrategias que mantengan el interés y el compromiso del estudiante a lo largo del tiempo. Esto se relaciona estrechamente con el seguimiento del progreso individual, que debe ser continuo y detallado para permitir ajustes oportunos en el proceso de aprendizaje.

La escalabilidad de la atención personalizada presenta un desafío particular en contextos educativos con recursos limitados, donde es necesario encontrar formas eficientes de proporcionar retroalimentación individualizada y apoyo personalizado a un gran número de estudiantes simultáneamente. Este desafío específico motiva la implementación de sistemas basados en [IA](#), particularmente aquellos que utilizan [RL](#) y arquitecturas [Transformers](#), que pueden proporcionar atención personalizada a escala mientras mantienen la calidad de la

instrucción.

### 3.1.7. Evaluación del Progreso

La evaluación efectiva del progreso en el aprendizaje de idiomas requiere un enfoque multidimensional y sistemático. Ellis (1994) enfatiza que la competencia comunicativa, que engloba tanto el conocimiento lingüístico como la capacidad de usarlo apropiadamente en contextos sociales, debe evaluarse a través de tareas que reflejen situaciones comunicativas auténticas.

La precisión gramatical, aunque no debe ser el único foco de evaluación, necesita ser monitoreada para asegurar que los estudiantes desarrollen un dominio adecuado de las estructuras lingüísticas fundamentales. Esta evaluación debe equilibrarse con la medición de la fluidez, que refleja la capacidad del estudiante para comunicarse de manera efectiva y natural en tiempo real.

Krashen (1982) sostiene que la comprensión auditiva y lectora requieren evaluaciones específicas que consideren diferentes tipos de textos y discursos, así como diversos propósitos comunicativos. Estas evaluaciones deben medir tanto la comprensión global como la capacidad de identificar detalles específicos.

La evaluación del progreso y la retroalimentación contextual son elementos cruciales que pueden beneficiarse significativamente de la integración de tecnologías avanzadas. Los sistemas basados en LLM y tecnologías de TTS y STT pueden proporcionar evaluaciones más precisas y detalladas de las habilidades lingüísticas del estudiante. Estos sistemas pueden analizar patrones de error, identificar áreas de mejora y proporcionar retroalimentación personalizada en tiempo real, superando las limitaciones de los métodos tradicionales de evaluación.

## 3.2. Inteligencia Artificial en Educación

La integración de la IA en el ámbito educativo ha transformado fundamentalmente la manera en que se concibe y se implementa el proceso de enseñanza-aprendizaje. Esta sección explora la evolución y el estado actual de los sistemas educativos inteligentes, con especial énfasis en su aplicación en la enseñanza de idiomas.

### 3.2.1. Evolución de los Sistemas de Aprendizaje Adaptativo

Los sistemas de aprendizaje adaptativo han evolucionado significativamente desde los primeros Intelligent Tutoring System (ITS) de la década de 1970. VanLehn VanLehn (2011) señala que esta evolución ha pasado por tres generaciones principales: sistemas basados en reglas, sistemas basados en el conocimiento del dominio, y sistemas adaptativos modernos que utilizan técnicas de aprendizaje automático y IA.

La primera generación se caracterizó por sistemas que seguían reglas predefinidas para adaptar el contenido. La segunda generación incorporó modelos del dominio más sofisticados

y comenzó a considerar el estado cognitivo del estudiante. La generación actual utiliza técnicas avanzadas de IA para crear experiencias de aprendizaje verdaderamente personalizadas, capaces de adaptarse en tiempo real a las necesidades y el progreso del estudiante.

### 3.2.2. Arquitecturas de Sistemas Educativos Inteligentes

Los sistemas educativos inteligentes modernos se construyen sobre arquitecturas modulares que integran múltiples componentes especializados. Anderson y Boyle (2020) identifican cuatro componentes principales:

1. El módulo experto contiene el conocimiento del dominio y las reglas pedagógicas que guían la instrucción.
2. El módulo del estudiante mantiene un modelo actualizado del conocimiento y las habilidades del aprendiz.
3. El módulo pedagógico determina las estrategias de enseñanza más apropiadas basándose en la información de los otros módulos.
4. La interfaz de usuario facilita la interacción entre el sistema y el estudiante.

### 3.2.3. Personalización y Adaptación Dinámica

La personalización y adaptación dinámica representan el núcleo de los sistemas educativos inteligentes modernos. Roll y Wylie (2018) describen cómo estos sistemas utilizan técnicas avanzadas de IA para:

- Construir y mantener modelos detallados del conocimiento del estudiante, incluyendo mapas de competencias, patrones de errores frecuentes y estilos de aprendizaje preferidos.
- Adaptar el contenido y el ritmo de instrucción en tiempo real, considerando tanto el rendimiento actual como el histórico del estudiante, y ajustando la dificultad de manera dinámica.
- Proporcionar retroalimentación personalizada que no solo identifique errores sino que ofrezca explicaciones contextuales y sugerencias específicas para la mejora.
- Identificar y abordar proactivamente áreas de dificultad mediante la predicción de posibles obstáculos en el aprendizaje.

### 3.2.4. Métodos de Evaluación Automática

Los métodos de evaluación automática han evolucionado significativamente con la integración de técnicas de NLP y IA. Baker e Inventado Baker y Inventado (2014) destacan la importancia de:



- Evaluación continua del progreso del estudiante mediante el análisis de múltiples indicadores de rendimiento, incluyendo precisión, velocidad de respuesta y patrones de interacción
- Análisis automático de patrones de error utilizando técnicas de [Data Mining](#) para identificar errores sistemáticos y conceptuales.
- Identificación temprana de dificultades de aprendizaje a través del monitoreo de métricas clave y la detección de desviaciones significativas en el rendimiento esperado.
- Generación de retroalimentación específica y constructiva utilizando técnicas de [NLP](#) para proporcionar explicaciones contextualizadas y sugerencias de mejora personalizadas.
- Adaptación dinámica de evaluaciones basada en el nivel demostrado por el estudiante, asegurando un balance óptimo entre desafío y apoyo.

### 3.2.5. Sistemas de Recomendación Educativa

Los [Sistema de Recomendación](#) en educación juegan un papel crucial en la personalización del aprendizaje. Estos sistemas utilizan técnicas de filtrado colaborativo y basado en contenido para:

- Recomendar rutas de aprendizaje personalizadas que consideren tanto el nivel actual como la velocidad de progreso del estudiante, adaptando dinámicamente la secuencia de contenidos para optimizar el proceso de aprendizaje.
- Adaptar el nivel de dificultad según el progreso del estudiante, utilizando algoritmos que analizan patrones de rendimiento para mantener un equilibrio óptimo entre desafío y motivación, evitando tanto la frustración como el aburrimiento.
- Identificar actividades complementarias apropiadas que refuercen áreas específicas de debilidad, proporcionando ejercicios adicionales y materiales de práctica focalizados en las necesidades individuales del estudiante.

La efectividad de estos sistemas depende en gran medida de su capacidad para equilibrar la exploración de nuevo contenido con la consolidación del aprendizaje existente, un desafío que se aborda mediante técnicas avanzadas de [RL](#), que permiten a los sistemas aprender y adaptarse continuamente a las necesidades cambiantes de los estudiantes.

## 3.3. Procesamiento del Lenguaje Natural y LLMs

El campo del [NLP](#) ha experimentado avances significativos en los últimos años, transformando fundamentalmente la manera en que interactuamos con el lenguaje natural. Esta sección examina las tecnologías clave que posibilitan sistemas educativos inteligentes para el aprendizaje de idiomas.

### 3.3.1. Arquitectura Transformer

La arquitectura Transformer, introducida por [Vaswani et al. \(2017\)](#), revolucionó el campo del [NLP](#) al proponer un modelo basado enteramente en mecanismos de atención. El componente fundamental de esta arquitectura es el [Mecanismo de Atención](#), que permite al modelo procesar secuencias de texto considerando las relaciones entre todas las palabras simultáneamente, superando las limitaciones de los modelos recurrentes tradicionales.

La arquitectura se compone de varios elementos clave:

- **Codificador-Decodificador:** El modelo utiliza una estructura de codificador-decodificador donde cada componente está compuesto por capas de [Auto-Atención](#) y redes [Feed-Forward](#). Esta estructura permite al modelo procesar texto de entrada y generar texto de salida de manera eficiente.
- **Atención Multi-Cabeza:** El mecanismo de atención multi-cabeza permite al modelo atender simultáneamente a diferentes aspectos de la entrada, capturando relaciones semánticas y sintácticas complejas. Cada cabeza de atención puede especializarse en diferentes tipos de relaciones lingüísticas.

### 3.3.2. Large Language Models (LLMs)

Los [LLM](#) representan la evolución más reciente en el procesamiento del lenguaje natural. [Brown et al. \(2020\)](#) demostró que estos modelos, entrenados en grandes cantidades de texto, pueden exhibir capacidades sorprendentes en una variedad de tareas lingüísticas. Las características principales de los LLMs incluyen:

Los [LLM](#) modernos se basan en arquitecturas [Transformers](#) con billones de parámetros, lo que les permite capturar patrones lingüísticos complejos y conocimiento del mundo real. El escalamiento en términos de parámetros y datos de entrenamiento ha demostrado mejorar continuamente el rendimiento en diversas tareas.

Una característica distintiva de los [LLM](#) es su capacidad de adaptar su comportamiento a nuevas tareas con pocos ejemplos, sin necesidad de reentrenamiento. Esta capacidad se manifiesta de tres formas principales:

- **Zero-shot learning:** El modelo puede realizar tareas sin ejemplos previos, basándose únicamente en instrucciones en lenguaje natural.
- **One-shot learning:** El modelo aprende de un único ejemplo para adaptar su comportamiento a una nueva tarea.
- **Few-shot learning:** El modelo utiliza varios ejemplos (típicamente 2-5) para comprender mejor el patrón o tarea requerida y mejorar su desempeño.

Esta flexibilidad en el aprendizaje en contexto es particularmente valiosa en entornos educativos, donde los modelos necesitan adaptarse rápidamente a diferentes estilos de enseñanza y necesidades específicas de los estudiantes.

### 3.3.3. Sistemas de Recuperación Aumentada con Generación (RAG)

Los sistemas **RAG**, introducidos por [Lewis et al. \(2020\)](#), combinan la capacidad generativa de los **LLM** con la recuperación de información específica. Esta arquitectura es particularmente relevante para aplicaciones educativas debido a sus características fundamentales.

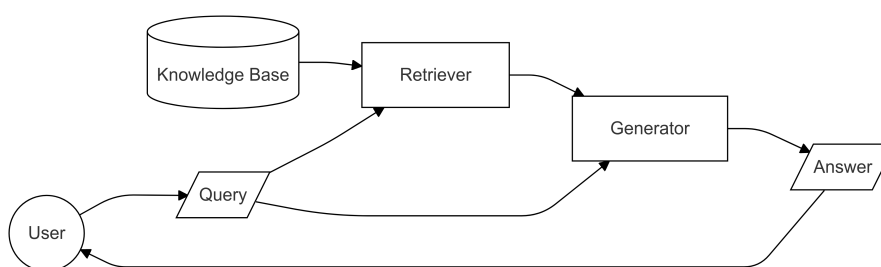
La combinación de generación de texto con recuperación de información permite respuestas más precisas y coherentes, ancladas en fuentes confiables. Además, estos sistemas pueden adaptarse a diferentes dominios de conocimiento mediante la actualización de la **Base de Conocimiento** subyacente, lo que los hace ideales para aplicaciones educativas que requieren contenido actualizado y relevante.

Otra ventaja significativa es que la capacidad de citar fuentes y materiales relevantes permite a los sistemas **RAG** personalizar el contenido educativo para cada estudiante, proporcionando referencias verificables y adaptando la información a las necesidades individuales.

#### 3.3.3.1. Arquitectura RAG

El sistema consta de tres componentes principales:

- Una **Base de Conocimiento** que almacena información estructurada y documentos relevantes para el dominio de aplicación
- Un **Recuperador** que accede a la base de conocimiento, utilizando técnicas avanzadas de indexación y búsqueda semántica para identificar la información más relevante
- Un **Generador** basado en **LLM** que produce respuestas considerando tanto el contexto como la información recuperada, asegurando coherencia y precisión en las respuestas



**Figura 3.1:** Flujo de información en un sistema RAG

El proceso de generación sigue tres pasos fundamentales:

1. **Recuperación de documentos relevantes:** El sistema vectoriza la consulta del usuario y busca en la base de conocimiento utilizando índices semánticos para encontrar documentos relacionados.
2. **Análisis y ranking de documentos:** Se evalúa la relevancia de los documentos recuperados considerando su similitud semántica con la consulta y la confiabilidad de las fuentes.

3. **Generación de respuestas:** El **LLM** integra el conocimiento recuperado con el contexto de la consulta para producir una respuesta coherente y precisa.

### 3.3.4. Aplicaciones y Ventajas de RAG en Educación

Los sistemas **RAG** ofrecen beneficios significativos para aplicaciones educativas, especialmente en la enseñanza de idiomas. Las ventajas principales incluyen:

- **Precisión y Confiabilidad:** Mayor precisión en la información proporcionada al combinar conocimiento estructurado con la flexibilidad de los **LLM**, reduciendo **Alucinaciones** y respuestas incorrectas al anclar la generación en fuentes confiables.
- **Trazabilidad y Verificabilidad:** Capacidad de citar fuentes y materiales relevantes, proporcionando referencias verificables para el contenido educativo.
- **Adaptabilidad y Actualización:** estos sistemas ofrecen adaptabilidad a diferentes dominios mediante la actualización de la base de conocimiento. Esto permite una actualización dinámica del contenido sin necesidad de reentrenar el modelo completo. Además, facilita la personalización del contenido educativo mediante la selección específica de fuentes relevantes para cada estudiante.

## 3.4. Aprendizaje por Refuerzo

El Aprendizaje por Refuerzo (**RL**) es una rama de la **IA** que se enfoca en desarrollar agentes inteligentes capaces de aprender a través de la interacción con un entorno. Esta sección explora los fundamentos teóricos del **RL** y su aplicación en el aprendizaje de idiomas.

### 3.4.1. Fundamentos Teóricos del RL

El Aprendizaje por Refuerzo proporciona un marco matemático ideal para la personalización del aprendizaje de idiomas. Basado en **Markov Decision Process (MDP)**, permite modelar el proceso de aprendizaje como una serie de decisiones secuenciales, donde el sistema debe seleccionar las actividades y contenidos más apropiados según el nivel y progreso del estudiante **Williams y Chen (2017)**.

En nuestro contexto, el estado representa el perfil actual del estudiante, incluyendo su dominio del idioma en diferentes áreas (comprensión, producción, vocabulario, gramática), mientras que las acciones corresponden a las diferentes intervenciones pedagógicas disponibles.

### 3.4.2. Proximal Policy Optimization (PPO)

PPO **Schulman et al. (2017)** es un algoritmo de **RL** que destaca por su estabilidad y eficiencia en el aprendizaje de políticas. En nuestro sistema de aprendizaje de idiomas, PPO se utiliza para optimizar la selección de actividades y la adaptación del contenido.

**3.4.2.1. Formulación Matemática**

El objetivo de PPO es maximizar la siguiente función objetivo:

$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t[\text{mín}(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)] \quad (3.1)$$

Donde:

- $r_t(\theta)$  es el ratio de probabilidades entre la política nueva y antigua
- $\hat{A}_t$  es la estimación de la ventaja
- $\epsilon$  es el parámetro de clipping (típicamente 0.2)

---

**Algoritmo 1:** Algoritmo *Proximal Policy Optimization* (PPO)

---

1. Inicializar los parámetros de la política  $\theta$  y el valor función  $\phi$
  2. Para cada iteración:
    - a) Recopilar conjunto de trayectorias  $\mathcal{D}_k = \{\tau_i\}$  ejecutando la política  $\pi_\theta$  en el entorno
    - b) Calcular ventajas estimadas  $\hat{A}_t$  usando función de valor actual  $V_\phi$
    - c) Para cada época de optimización:
      - 1) Calcular ratio de probabilidad  $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$
      - 2) Calcular pérdida recortada:
$$L^{CLIP}(\theta) = \hat{\mathbb{E}}_t[\text{mín}(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)]$$
      - 3) Actualizar  $\theta$  minimizando  $-L^{CLIP}(\theta)$  usando descenso de gradiente
      - 4) Actualizar función de valor  $\phi$  minimizando error cuadrático medio
    - d) Actualizar  $\theta_{old} \leftarrow \theta$
  3. Devolver la política optimizada  $\pi_\theta$
- 

**3.4.2.2. Aplicación en el Sistema**

En nuestro contexto educativo:

- **Estado ( $S$ ):** Representa el perfil actual del estudiante.

$$S = \{\text{vocabulario\_nivel} = B1, \text{gramática\_nivel} = A2, \text{pronunciación\_nivel} = B2\} \quad (3.2)$$

- **Acciones ( $\mathcal{A}$ ):** Selección de actividades y sus parámetros.

$$\mathcal{A} = \{\text{ejercicio\_gramática\_A2}, \text{práctica\_vocabulario\_B1}, \text{diálogo\_pronunciación\_B2}\} \quad (3.3)$$

- **Recompensa ( $\mathcal{R}$ ):** Evalúa el éxito de cada acción. Por ejemplo, si después de un ejercicio de gramática el estudiante mejora su precisión del 60 % al 80 %,  $\mathcal{R} = +20$
- **Política ( $\pi$ ):** Determina qué acción tomar en cada estado. Por ejemplo, si el estudiante muestra consistentemente errores en gramática,  $\pi$  seleccionará más ejercicios gramaticales

### 3.4.2.3. Sistema de Recompensas

La [Función de Recompensa](#) se diseña específicamente para el aprendizaje de idiomas, evaluando el desempeño y proporcionando retroalimentación a través de múltiples dimensiones:

- **Recompensas inmediatas:** Incluyen la precisión en las respuestas y ejercicios, mejora en la pronunciación y fluidez, uso correcto de estructuras gramaticales, y la adquisición y retención de vocabulario.
- **Recompensas a largo plazo:** Consideran el progreso sostenido en múltiples dimensiones lingüísticas, la mejora en la competencia comunicativa general, y la retención y aplicación de conocimientos previos.
- **Ajustes dinámicos:** Comprenden la calibración automática de pesos de recompensa, adaptación a diferentes estilos y velocidades de aprendizaje, y el balanceo entre diferentes competencias lingüísticas.

### 3.4.3. Evaluación de Políticas de Aprendizaje

La evaluación de la [Política](#) en sistemas de aprendizaje de idiomas requiere un enfoque multidimensional que considere tanto aspectos cuantitativos como cualitativos. [Williams y Chen \(2017\)](#) propone un marco de evaluación que examina:

- **Progreso en competencias lingüísticas específicas:** Incluye la mejora en precisión gramatical y uso de estructuras, expansión del vocabulario activo y pasivo, desarrollo de fluidez y pronunciación, y avance en comprensión auditiva y lectora.
- **Efectividad de la personalización:** Abarca la adaptación a estilos individuales de aprendizaje, respuesta a patrones de error específicos, ajuste dinámico del nivel de dificultad y personalización de contenido temático.
- **Eficiencia en el tiempo de aprendizaje:** Considera la tasa de adquisición de nuevos conceptos, reducción en tiempo de dominio de habilidades, optimización de intervalos de repaso y minimización de redundancia en ejercicios.

- **Engagement y retención del estudiante:** Evalúa los niveles de participación activa, tasas de completación de actividades, persistencia en el programa de aprendizaje y satisfacción reportada por el estudiante.

La evaluación se realiza mediante métricas cuantitativas específicas:

$$\text{Efectividad} = \frac{\text{Objetivos Alcanzados}}{\text{Tiempo Invertido}} \times \text{Factor de Dificultad} \quad (3.4)$$

$$\text{Índice de Personalización} = \frac{\sum_{i=1}^n \text{Adaptaciones Exitosas}_i}{n} \times \text{Tasa de Progreso} \quad (3.5)$$

Estas métricas se complementan con análisis cualitativo continuo y retroalimentación directa de los estudiantes para asegurar una evaluación holística de la [Política](#).

### 3.5. Tecnologías de Procesamiento de Voz

El procesamiento de voz en sistemas de aprendizaje de idiomas involucra dos procesos fundamentales: el reconocimiento automático del habla ([STT](#)) y la síntesis de voz ([TTS](#)). Estos procesos representan transformaciones complementarias entre el dominio acústico y el lingüístico.

#### 3.5.1. Reconocimiento Automático del Habla (STT)

El proceso de STT transforma señales acústicas en texto, involucrando múltiples etapas de procesamiento y análisis. Este proceso se fundamenta en principios de procesamiento de señales y modelos probabilísticos del lenguaje [Graves et al. \(2013\)](#).

##### 3.5.1.1. Procesamiento de la Señal Acústica

- **Preprocesamiento Acústico:** La señal de audio cruda se somete a técnicas de reducción de ruido, normalización de amplitud y segmentación en tramas. Este proceso mejora la calidad de la señal y la prepara para el análisis posterior.
- **Extracción de Características:** Se extraen representaciones espectrales como coeficientes [Mel-Frequency Cepstral Coefficients \(MFCC\)](#), que capturan las características acústicas relevantes para el reconocimiento del habla.
- **Normalización de Características:** Las características extraídas se normalizan para reducir variaciones no lingüísticas como diferencias en el volumen o el canal de grabación.

### 3.5.1.2. Proceso de Reconocimiento

- **Modelado Acústico:** Se analiza la relación entre las características acústicas y las unidades fonéticas del habla, considerando variaciones en pronunciación y contexto fonético.
- **Modelado del Lenguaje:** Se incorpora conocimiento sobre la estructura del lenguaje, incluyendo probabilidades de secuencias de palabras y restricciones gramaticales.
- **Decodificación:** Se combina la información acústica y lingüística para determinar la secuencia más probable de palabras, utilizando algoritmos de búsqueda como [Viterbi](#) o [Beam Search](#).

### 3.5.2. Síntesis de Voz (TTS)

La síntesis de voz realiza la transformación inversa, convirtiendo texto en señales de habla mediante un proceso que combina análisis lingüístico y generación de señales acústicas [Taylor \(2009\)](#).

#### 3.5.2.1. Procesamiento Lingüístico

- **Análisis de Texto:** Se procesa el texto de entrada para identificar su estructura lingüística, incluyendo tokenización, normalización y análisis sintáctico.
- **Conversión Grafema-Fonema:** Se transforma el texto escrito en su representación fonética, considerando reglas de pronunciación y excepciones específicas del idioma.
- **Análisis Prosódico:** Se determinan patrones de entonación, duración y énfasis basados en la estructura sintáctica y semántica del texto.

#### 3.5.2.2. Generación de Voz

- **Modelado Prosódico:** Se generan patrones detallados de pitch, duración y energía para cada fonema, considerando el contexto lingüístico y emocional.
- **Generación de Características Acústicas:** Se producen representaciones espectrales intermedias que codifican las propiedades acústicas deseadas del habla.
- **Síntesis de Forma de Onda:** Se genera la señal de audio final mediante técnicas de síntesis que pueden ser concatenativas, paramétricas o basadas en modelos neuronales.

### 3.5.3. Integración en Sistemas de Aprendizaje

La combinación de STT y TTS en sistemas educativos permite crear ciclos completos de interacción oral:



- **Ciclo de Retroalimentación:** El sistema puede generar ejemplos de pronunciación (TTS), analizar la producción del estudiante (STT) y proporcionar retroalimentación específica.
- **Análisis de Precisión:** La comparación entre la transcripción del habla del estudiante y el texto objetivo permite evaluar la precisión de pronunciación y fluidez.
- **Adaptación Dinámica:** Los resultados del análisis permiten ajustar parámetros como velocidad del habla, complejidad del contenido y umbral de aceptación de pronunciación.

La integración de estas tecnologías en sistemas de aprendizaje de idiomas permite una interacción más natural y efectiva, facilitando la práctica oral y la mejora de la competencia comunicativa de los estudiantes.

**Material**

4

# Métodos

5

# Resultados

6

# Conclusiones

7



# Glosario

**Alucinaciones** Errores en la generación de texto que resultan en respuestas incoherentes o incorrectas.

**Auto-Atención** Mecanismo que permite a un modelo evaluar las relaciones entre todas las posiciones de una secuencia.

**Base de Conocimiento** Conjunto de datos estructurados que almacena información relevante para un sistema de recuperación de información.

**Beam Search** Algoritmo de búsqueda heurística que explora un grafo construyendo el grafo gradualmente desde la raíz, expandiendo el nodo más prometedor en un conjunto limitado de nodos.

**Código Abierto** Software cuyo código fuente está disponible públicamente y puede ser modificado y distribuido por cualquier persona.

**Data Mining** Proceso de descubrir patrones y relaciones en grandes conjuntos de datos.

**Feed-Forward** Capa de red neuronal que aplica una transformación lineal seguida de una función de activación.

**Función de Recompensa** Función que define la retroalimentación que recibe el agente basada en el progreso del estudiante.

**Generador** Componente de un sistema de recuperación de información que crea respuestas a partir de los documentos recuperados.

**IA** Inteligencia Artificial - Conjunto de tecnologías que permiten a las máquinas aprender, razonar y tomar decisiones.

**ITS** Intelligent Tutoring System - Sistema de Tutoría Inteligente.

**LLM** Large Language Model - Modelo de lenguaje de gran escala.

**MDP** Markov Decision Process - Marco matemático para modelar la toma de decisiones en situaciones donde los resultados son parcialmente aleatorios y parcialmente bajo control.

**Mecanismo de Atención** Componente clave de la arquitectura Transformer que permite al modelo enfocarse en diferentes partes de la entrada según su relevancia.

**MFCC** Mel-Frequency Cepstral Coefficients. Coeficientes que representan el espectro de potencia a corto plazo de un sonido, basados en una transformación coseno lineal de un espectro de potencia logarítmico en una escala de frecuencia mel no lineal.

**NLP** Natural Language Processing - Procesamiento del Lenguaje Natural.

**Política** Estrategia que sigue un agente para determinar sus acciones basándose en el estado actual del estudiante.

**PyTorch** Biblioteca de aprendizaje profundo de código abierto desarrollada por Meta.

**RAG** Retrieval-Augmented Generation - Sistema que combina la recuperación de información con la generación de texto.

**Recuperador** Componente de un sistema de recuperación de información que selecciona documentos relevantes a partir de una consulta.

**RL** Reinforcement Learning - Aprendizaje por Refuerzo, una rama de la IA que permite a los sistemas aprender a través de la interacción con un entorno.

**Sistema de Recomendación** Sistema que sugiere contenido relevante basado en el perfil del usuario y su comportamiento.

**Sistema Multi-Agente** Sistema compuesto por múltiples agentes inteligentes que interactúan entre sí para resolver problemas complejos.

**SLA** Second Language Acquisition - Proceso de adquisición de una segunda lengua.

**STT** Speech-to-Text (Voz a Texto). Sistema que convierte el habla en texto escrito mediante reconocimiento automático del habla.

**TensorFlow** Biblioteca de código abierto para aprendizaje automático desarrollada por Google.

**Transformers** Arquitectura de red neuronal que ha revolucionado el procesamiento del lenguaje natural.

**TTS** Text-to-Speech (Texto a Voz). Sistema que convierte texto escrito en habla sintetizada.



**Viterbi** Algoritmo que encuentra la secuencia más probable de estados ocultos en un modelo oculto de Markov, comúnmente usado en reconocimiento de voz para decodificación.

## Apéndice A



# Bibliografía

- Anderson, J. R. y Boyle, C. F. (2020). Adaptive learning systems in modern education. *Journal of Computer Assisted Learning*.
- Baker, R. S. y Inventado, P. S. (2014). Educational data mining and learning analytics. *Learning Analytics*.
- Brown, T. et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*.
- Coyle, D., Hood, P., y Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.
- Ellis, R. (1994). The study of second language acquisition. *Oxford University Press*.
- Fries, C. C. (1945). *Teaching and learning English as a foreign language*. University of Michigan Press.
- Gouin, F. (1892). *The art of teaching and studying languages*. Heath, D.C.
- Graves, A., Mohamed, A.-r., y Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *IEEE International Conference on Acoustics, Speech and Signal Processing*.
- Hymes, D. (1972). *On communicative competence*. University of Pennsylvania Press.
- Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Lewis, P. et al. (2020). Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*.
- Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge University Press.
- Richards, J. C. y Rodgers, T. S. (2000). *Approaches and methods in language teaching*. Cambridge University Press.
- Roll, I. y Wylie, R. (2018). Learning analytics and ai: Politics, pedagogy and practices. *British Journal of Educational Technology*.
- Schulman, J. et al. (2017). Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- Taylor, P. (2009). Text-to-speech synthesis. *Cambridge university press*.
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*.
- Vaswani, A. et al. (2017). Attention is all you need. *Advances in neural information processing systems*.
- Williams, R. y Chen, D. (2017). The use of reinforcement learning algorithms in adaptive education. *Journal of Educational AI*.