

---

This is the **published version** of the master thesis:

Lu, Wang; Vargas-Urpi, Mireia, dir.; Casas-Tost, Helena, coord. El uso de ChatGPT y Gemini para la preparación de actividades de comprensión oral de CLE : un estudio comparativo. 2024. 117 pag. (Màster Universitari en Didàctica del Xinès per a Hispanoparlants)

---

This version is available at <https://ddd.uab.cat/record/300353>

under the terms of the  license

**UNIVERSITAT AUTÒNOMA DE BARCELONA**

Máster en Didáctica del Chino para Hispanohablantes

**El uso de ChatGPT y Gemini para la preparación  
de actividades de comprensión oral de CLE: un  
estudio comparativo**

**使用 ChatGPT 和 Gemini 准备 CLE 听力材料：比  
较研究**

Autora: Wang Lu (王露)

Tutora: Dra. Mireia Vargas-Urpí

Coordinadora: Dra. Helena Casas-Tost

18 de julio, 2024

## **AGRADECIMIENTOS**

Hasta aquí, he completado mi tesina del máster. Estoy profundamente agradecida a mi tutora, Dra. Mireia Vargas-Urpí, por su detallada orientación durante todo el proceso de investigación. Siempre revisó mi trabajo a tiempo y me proporcionó comentarios constructivos. Sin su aliento y ayuda, no habría podido terminar esta tesina.

En segundo lugar, me gustaría agradecer a todos mis profesores de la Universidad Autónoma de Barcelona. Gracias por los conocimientos compartidos a través de clases interesantes y meticulosamente diseñadas. Esto ha ampliado mi perspectiva y me ha permitido disfrutar del proceso de aprendizaje.

En tercer lugar, quiero expresar mi gratitud a mis padres, quienes siempre me brindaron su cariño y apoyo. Aunque estamos separados por miles de kilómetros, siempre he sentido su cuidado y preocupación. También quiero agradecer a mi novio. Durante mi tiempo de estudio, enfrenté muchas dificultades y momentos de incertidumbre, y él siempre estuvo allí para apoyarme y animarme a seguir adelante. Muchas gracias a todos mis familiares y amigos quienes siempre me han amado y han permitido que crezca libremente bajo su amor.

Finalmente, quiero agradecerme a mí misma. Después de tres años desde mi graduación del grado, aún tuve el coraje de perseguir y realizar mi sueño de estudiar el máster aquí. Espero que en el futuro pueda mantener siempre esta valentía y pasión en mi vida.

¡Muchísimas gracias a todos!

Wang Lu

Barcelona, 18 de julio de 2024

## RESUMEN

En el contexto de la integración de las herramientas de inteligencia artificial en el aula, la selección y aplicación de estas para facilitar la enseñanza representa un desafío considerable para los profesores que están acostumbrados a los modelos tradicionales de enseñanza. En este sentido, este trabajo tiene como objetivo explorar ChatGPT y Gemini para determinar cuál es más útil, en otras palabras, para identificar cuál puede reducir más la carga de trabajo al docente en el aula de chino como lengua extranjera (CLE), específicamente en la preparación de actividades de comprensión oral. Por un lado, nos hemos concentrado en la evaluación y comparación de los resultados generados por ChatGPT y Gemini. Para ello, hemos hecho tres rúbricas para analizar tres aspectos: los textos sobre los que se hará la comprensión oral, los ejercicios (las preguntas) para valorar la comprensión oral y, por último, la dicción oral. Por otro lado, también hemos analizado el proceso hasta conseguir las actividades, es decir, después de introducir el primer comando (prompt), hemos observado si necesitamos introducir una nueva instrucción para optimizar el resultado generado. Según los resultados del estudio, Gemini y ChatGPT tienen sus propias ventajas a la hora de generar materiales de comprensión oral. Gemini es más útil en generar los textos, pero ChatGPT funciona mejor en el diseño de los ejercicios y es más eficiente para conseguir los contenidos que deseamos. Con este trabajo, esperamos dar consejos y propuestas al docente para mejorar su enseñanza con el uso de herramientas de IA.

**PALABRAS CLAVE:** ChatGPT, Gemini, actividades de comprensión oral, chino como lengua extranjera, sistemas de Inteligencia Artificial Generativa

**KEY WORDS:** ChatGPT, Gemini, listening comprehension activities, Chinese as a foreign language, Generative Artificial Intelligence systems

## 摘要

在人工智能工具融入课堂的背景下，如何选择和应用这些工具来促进教学，对于习惯于传统教学模式的教师来说是一个巨大的挑战。基于此现状，本文旨在探索 ChatGPT 和 Gemini 哪一款软件在对外汉语课堂，尤其是在听力理解练习的准备上更有效用，即更能减轻教师的工作。一方面，我们聚焦于评估与比较 ChatGPT 和 Gemini 生成的结果。为此，我们制作了三个评分标准，分别从听力文本、听力题目和语音三个方面进行分析。另一方面，我们还分析了每个工具达到我们所期望的结果需经历的过程，即在输入第一个指令后，我们是否还需要输入新的指令对 ChatGPT 和 Gemini 生成的结果进行优化。研究表明，Gemini 和 ChatGPT 在生成汉语听力材料方面各有千秋。Gemini 在生成文本方面表现更好，而 ChatGPT 在设计练习方面更有优势，同时可以更高效地生成令人满意的结果。通过这项研究，我们希望为教师提供意见和建议，以便他们更好地使用人工智能工具进行教学。

**关键词：**ChatGPT，Gemini，听力理解练习，对外汉语，生成式人工智能系统

## ÍNDICE:

RESUMEN .....	2
摘要 .....	3
1. INTRODUCCIÓN .....	6
2. MARCO CONTEXTUAL: la Inteligencia Artificial Generativa .....	9
2.1 Descripción del estado de las herramientas de IA en general .....	9
2.1.1. ¿Qué son los large language models? .....	9
2.1.2 ¿Qué es la <i>natural language generation</i> ? .....	11
2.1.3 ¿Qué es ChatGPT y qué funcionalidades tiene? .....	12
2.1.4 ¿Qué es Gemini y qué funcionalidades tiene? .....	16
2.1.5 ¿Qué diferencias hay entre ChatGPT y Gemini? .....	16
2.2 Estudios o experiencias previas en la aplicación de las herramientas de IA en enseñanza de lenguas .....	20
3. MARCO TEÓRICO: la comprensión oral en CLE .....	23
3.1 La recomendación del Marco Común Europeo de Referencia para las Lenguas .....	23
3.2 La importancia de la comprensión oral en el HSK .....	27
3.3 Estudios previos sobre la enseñanza de comprensión oral en CLE .....	30
4. METODOLOGÍA .....	34
4.1 Recogida de datos .....	34
4.2 Análisis de los datos .....	38
4.2.1 Rúbrica para evaluar los textos de la comprensión oral generados por ChatGPT y Gemini .....	39
4.2.2 Rúbrica para evaluar los ejercicios de la comprensión oral generados por ChatGPT y Gemini .....	43
4.2.3 Rúbrica para evaluar la dicción oral de la comprensión oral generados por ChatGPT y Gemini .....	46
5. ANÁLISIS DE LOS MATERIALES .....	49
5.1 Análisis de ChatGPT .....	49
5.1.1 Análisis de los textos y ejercicios de verdadero y falso generado por ChatGPT 3.5 (Ejemplo 1) .....	50

5.1.2 Análisis de los textos y ejercicios de elección múltiple generado por ChatGPT 3.5 (Ejemplo 2).....	54
5.1.3 Análisis de los textos y ejercicios de elección múltiple generado por ChatGPT 3.5 (Ejemplo 3).....	58
5.1.4 Análisis de la dicción oral generada por ChatGPT 3.5.....	64
5.2 Análisis de Gemini.....	67
5.2.1 Análisis de los textos y ejercicios de verdadero y falso generado por Gemini (Ejemplo 1).....	67
5.2.2 Análisis de los textos y ejercicios de elección múltiple generado por Gemini (Ejemplo 2).....	72
5.2.3 Análisis de los textos y ejercicios de elección múltiple generado por Gemini (Ejemplo 3).....	77
5.2.4 Análisis de la dicción oral generado por Gemini.....	84
5.3 Comparación entre ChatGPT y Gemini.....	86
5.3.1 Los textos.....	87
5.3.2 Los ejercicios.....	89
5.3.3 La dicción oral.....	92
6. ANÁLISIS DEL PROCESO.....	96
6.1 Análisis de ChatGPT.....	96
6.2 Análisis de Gemini.....	97
6.3 Comparación entre ChatGPT y Gemini.....	103
7. CONCLUSIONES.....	105
8. BIBLIOGRAFÍA.....	109

## 1. INTRODUCCIÓN

Hoy en día hay muchas investigaciones que sugieren que la Inteligencia Artificial puede utilizarse para la enseñanza y el aprendizaje de idiomas. Por ejemplo, los alumnos pueden chatear con un chatbot para simular entornos lingüísticos reales fuera de la clase, y los profesores pueden utilizar ChatGPT para revisar la escritura de los alumnos, redactar planes de clase y diseñar tareas o preguntas (Hong, 2023: 40-42). Comprendemos que la introducción de software de IA en la educación tiene perspectivas prometedoras.

Sin embargo, todavía falta investigación sobre cómo aplicar realmente herramientas de IA para preparar las actividades en el aula, para valorar qué efectividad tienen los materiales didácticos generados, y también para determinar qué problemas pueden surgir dado que algunos estudios dicen que ChatGPT tiene el potencial de generar información incorrecta o engañosa (Borji, 2023).

Por lo tanto, en este estudio exploraremos el uso de herramientas de IA como ChatGPT y Gemini en la preparación de actividades de comprensión oral en el aula de chino como lengua extranjera. A través de este estudio, esperamos proporcionar a los profesores de CLE métodos eficaces e instrucciones útiles para utilizar ChatGPT y Gemini, así como ideas para preparar materiales de enseñanza. Al mismo tiempo, analizaremos y compararemos las ventajas y desventajas de estas dos herramientas a la hora de generar contenidos, con el fin de ayudar al docente a elegir un software más adecuado para su uso práctico, y a aumentar lo largo y evitar lo corto.

En resumen, nuestro objetivo es explorar cómo ChatGPT y Gemini pueden ser utilizados para generar materiales de comprensión oral en clases de chino como lengua extranjera. Concretamente, este objetivo se concreta en dos objetivos específicos:

1. Evaluar los materiales de comprensión oral creados por herramientas de IA (es decir, evaluar el producto final), incluyendo la adecuación del texto o diálogo, de las preguntas y de la dicción oral.

2. Determinar en qué medida es eficiente el uso de herramientas de IA como asistente del docente de chino (es decir, examinar el proceso hasta conseguir los ejercicios).

Con la evaluación y comparación de las herramientas, esperamos responder finalmente a la siguiente pregunta: ¿Cuál resultará más útil para generar materiales de comprensión oral en clases de CLE, ChatGPT o Gemini?

Esta tesina se divide en ocho capítulos:

En el primer capítulo, presentamos la motivación por este tema, la justificación de la elección del tema, los objetivos y una breve explicación de la estructura.

En el segundo capítulo, revisamos los estudios sobre la Inteligencia Artificial Generativa. Por un lado, describimos el estado de la IA en general, mencionamos conceptos como *large language models* y *natural language generation*, y comparamos las funcionalidades de ChatGPT y Gemini. Por otro lado, resumimos estudios y experiencias previas en la aplicación de herramientas de IA en la enseñanza de idiomas.

En el tercer capítulo, introducimos el marco teórico sobre la comprensión oral en CLE. En él, nos referimos a los requisitos para la comprensión oral en el *Marco Común Europeo de Referencia para las Lenguas* y el *Hanyu Shuiping Kaoshi* y hacemos un repaso de los estudios previos sobre la enseñanza de comprensión oral en CLE .

En el cuarto capítulo, planteamos la metodología del trabajo. Primero, diseñamos las rúbricas para evaluar textos, ejercicios y la dicción oral generados por ChatGPT 3.5 y Gemini. Luego, les pedimos a ChatGPT 3.5 y Gemini que generen material similar al de comprensión oral del Nuevo HSK Nivel 4. Después, evaluamos el contenido utilizando las rúbricas. Por último, comparamos los resultados de las evaluaciones.

En el quinto capítulo, mostramos los materiales de comprensión oral generados por ChatGPT 3.5 y Gemini, analizamos el contenido en detalle, evaluándolos y puntuándolos en términos de texto, ejercicios y dicción oral. Finalmente, integramos los resultados del análisis de los dos programas y los resumimos en forma de gráfica para hacer una comparación.

En el sexto capítulo, analizamos el proceso hasta conseguir los ejercicios (es decir, las preguntas de comprensión oral). En concreto, examinamos si es necesario añadir más instrucciones para optimizar el contenido inicial generado por ChatGPT y Gemini una vez que hemos introducido las instrucciones por primera vez, con el objetivo de determinar la eficiencia.

En el séptimo capítulo, resumimos el análisis y respondemos a la pregunta planteada inicialmente: ¿Cuál resultará más útil para generar materiales de comprensión oral en clases de CLE, ChatGPT o Gemini? Al mismo tiempo, se formulan algunas sugerencias y consideraciones prácticas para el uso de la IA por parte del docente.

El último capítulo contiene la bibliografía a la que nos referimos en el transcurso del estudio.

## **2. MARCO CONTEXTUAL: la Inteligencia Artificial Generativa**

En esta parte, se presentarán el marco contextual para esta investigación. En primer lugar, se describe el estado actual de las herramientas de inteligencia artificial (IA). Se explica la definición de los *large language models* (LLM), las herramientas de *natural language generation* (NLG) más utilizadas y sus principios de funcionamiento, así como las principales características y diferencias entre los objetos de estudio: ChatGPT y Gemini. Esto proporciona un marco contextual para situar el uso de ChatGPT y Gemini en la generación de ejercicios de comprensión oral.

En segundo lugar, se presenta una revisión de estudios o experimentos previos sobre la aplicación de herramientas de IA en la enseñanza de idiomas, lo que proporciona ideas para la metodología y el diseño de diferentes actividades de comprensión oral en este estudio.

### **2.1 Descripción del estado de las herramientas de IA en general**

En este apartado, se presentará el funcionamiento de los *large language models* (LLM), explorando sus aplicaciones en el campo educativo, así como las técnicas de *natural language processing* (NLP) en las que se basan, con un enfoque especial en la *natural language generation* (NLG), que es fundamental para la realización de esta investigación. Además, se comparará generalmente las funcionalidades de ChatGPT y Gemini, modelos basados en la tecnología de NLG, mediante la revisión de literatura y fuentes tecnológicas disponibles.

#### **2.1.1. ¿Qué son los large language models?**

Ante todo, nos conviene definir bien qué son los *large language models*. Tal como indican Yan Lixiang et al. (2024: 92), un LLM (*Large Language Model*) es un modelo de inteligencia artificial generativa entrenado con grandes volúmenes de datos textuales, diseñado para producir contenido escrito que se asemeje a la calidad humana a partir de instrucciones en lenguaje natural. Según Meyer et al. (2023), los datos textuales pueden ser libros, artículos y páginas web. También Esnaola y Ramón añaden más sobre sus funcionalidades (2023: 1): “Son capaces de, por ejemplo,

reconocer, resumir, traducir, predecir y generar texto.” Para llegar a una conclusión, se coincide con lo que resumen Hadi et al. (2023: 5-6), quienes definen los *large language models* como: “*a type of AI model that can process and generate natural language text. These models are typically trained on massive amounts of text data and use deep learning techniques to learn the patterns and structures of language.*”

En la actualidad, algunos de los *large language models* bien conocidos incluyen la serie de modelos GPT (*Generative Pre-trained Transformer*) de OpenAI, como GPT-3.5 y GPT-4. También se destacan modelos como PaLM de Google, Gemini y T5. Además, los modelos de la serie LLaMA de Meta han ganado popularidad (Esnaola y Ramón, 2023: 1, Naveed et al., 2023: 19). Herramientas como ChatGPT de OpenAI, Bing Chat de Microsoft y Gemini de Google se han popularizado, especialmente para uso conversacional, ofreciendo interfaces simplificadas para interactuar con algunos de estos modelos (Esnaola y Ramón, 2023: 1).

Los LLMs pueden ser aplicados en diversos ámbitos, como la medicina, la educación, la ciencia, las matemáticas, el derecho y el sector financiero (Naveed et al., 2023: 29). Como docentes, cabe prestar atención en la aplicación de LLMs en el ámbito educativo. Desde el punto de vista de Naveed et al. (2023: 29), los LLMs pueden reducir la carga de trabajo de profesores, debido a que son capaces de crear planes de lecciones, calificar tareas y generar contenido educativo diverso e inclusivo, lo que permita los educadores dedicar más tiempo a la enseñanza y la atención individualizada a los estudiantes. De acuerdo con Kasneci et al. (2023):

*“For lesson planning, large language models can also assist teachers in the creation of (inclusive) lesson plans and activities. Teachers can input to the models the corpus of document based on which they want to build a course. The output can be a course syllabus with short description of each topic. Moreover, they can be used to generate targeted and personalized practice problems and quizzes, which can help to ensure that students are mastering the material.”*

Por lo tanto, estos planteamientos sugieren que los LLMs ayudarán a los profesores a reducir su trabajo en la enseñanza, les proporcionan contenido creativo y contribuyen a mejorar la eficiencia.

### **2.1.2 ¿Qué es la *natural language generation*?**

Basándonos en la sección anterior, se infiere que ChatGPT y Gemini son *large language models*. En este sentido, se plantea la pregunta: ¿qué tecnologías emplean ChatGPT y Gemini para cumplir con su funcionalidad de la comprensión, predicción y generación de texto? Para ello, resulta necesario conocer la tecnología de *Natural Language Processing* (NLP) y *Natural Language Generation* (NLG).

*Natural Language Processing*, según lo que define Liddy (2001: 3), se trata de un conjunto de técnicas computacionales teóricamente motivadas para analizar y representar textos naturalmente con el propósito de lograr un procesamiento del lenguaje similar al humano para llevar a cabo una variedad de tareas. Hay dos aspectos importantes en la comunicación entre el humano y el ordenador: comprensión y generación (Paris, Swartout y Mann, 2013: 10).

La *Natural Language Generation* (NLG) es un subcampo de NLP que se centra más en la generación del lenguaje natural a partir de datos estructurados. Tal como dicen Vicente et al. (2015: 2), el objetivo de la NLG es investigar cómo se pueden crear aplicaciones informáticas capaces de producir por sí mismas textos de alta calidad en lenguaje natural. Afirma que los sistemas de NLG pueden generar diversos tipos de textos, como textos informativos, resúmenes (médicos, de ingeniería, financieros, deportivos, entre otros), textos simplificados, textos persuasivos, diálogos, explicaciones de razonamientos y recomendaciones. El chatbot ChatGPT se trata de una de las últimas herramientas basadas en tecnología de generación de lenguaje natural desarrollado por OpenAI (Franciscu, 2023). Además, también existen muchas otras herramientas de NLG como Claude, Google Bard AI (Gemini), Microsoft Bing Chat y ERNIE Bot.

### 2.1.3 ¿Qué es ChatGPT y qué funcionalidades tiene?

El avance revolucionario en el campo de la inteligencia artificial y el NLP ha propiciado la evolución de modelos de lenguaje progresivamente más sofisticados y especializados, conocidos como los LLMs. Entre los LLMs más influyentes, el *Generative Pre-trained Transformer* (GPT) desarrollado por OpenAI y su variante, ChatGPT, han surgido como una herramienta poderosa debido a sus capacidades de generación de lenguaje y a su amplio rango de aplicaciones, especialmente en contextos conversacionales (Nazir y Wang, 2023: 1).

Para definir el *Generative Pre-trained Transformer*, utilizaremos el resumen de *World Economic Forum* (2023), que dice que es un tipo de LLM que emplea técnicas de aprendizaje profundo para producir texto que se asemeja al generado por humanos. Se llamado “*Generative*” debido a su capacidad para generar nuevo texto basado en la información de entrada que reciben. El término “*Pre-trained*” se refiere a que estos modelos son entrenados previamente en un extenso conjunto de datos de texto antes de ser ajustados para tareas específicas. Además, se denominado “*Transformer*” debido a que utiliza una arquitectura de red neuronal basada en transformadores para procesar la información de entrada y generar el texto resultante.

Tal como concluyen Nazir y Wang (2023: 2) en su artículo, la evolución de GPT fue impulsada con el objetivo de desarrollar un modelo de lenguaje de IA altamente avanzado capaz de ayudar en una variedad de tareas, como análisis de datos, traducción de texto y generación de contenido escrito. Comenzando con GPT-1, la primera versión del modelo, demostró un buen desempeño en diversas áreas del procesamiento del lenguaje natural y generación de texto coherente. Luego se introdujo GPT-2, una versión mejorada que mostró una notable habilidad para producir texto coherente y relevante. GPT-3, lanzado en 2020, es la versión más avanzada y completa del modelo, con un gran número de parámetros. Este modelo destacó por su capacidad para comprender el lenguaje natural, traducir, generar texto y resolver problemas complejos de razonamiento. ChatGPT, basado en la arquitectura

de GPT-3.5, se diseñó especialmente para mantener conversaciones interactivas y dinámicas. En comparación con ChatGPT 3.5, ChatGPT 4 es más fiable, creativo y capaz de manejar instrucciones más detalladas. El 13 de mayo de 2024, OpenAI lanzó la versión GPT-4o, que está optimizada aún más sobre la base de 3.5 y 4.0. GPT-4o es capaz de razonar sobre audio, imágenes y texto en tiempo real. Puede observar directamente la entonación, varios hablantes o el ruido de fondo, y también puede emitir risas, cantar o expresar emociones.

Dentro del mismo trabajo, Nazir y Wang (2023: 4) también nos han enumerado las funcionalidades de ChatGPT y sus aplicaciones posibles, algunas de las cuales son las siguientes:

**Asistentes personales:** Los usuarios pueden interactuar con ChatGPT para realizar diversas acciones, como programar citas, establecer recordatorios, leer noticias, controlar dispositivos domésticos inteligentes y enviar mensajes. ChatGPT puede proporcionar rápidamente explicaciones sobre una amplia gama de temas, encontrar respuestas a las preguntas de los usuarios y ofrecer acceso oportuno a información relevante. Muchos estudiantes utilizan ChatGPT en combinación con Grammarly, Wikipedia, Google translate, Quillbot, etc.

**Generación de contenido:** ChatGPT destaca en la redacción de informes, correos electrónicos, materiales de marketing y gestión de publicaciones en redes sociales. Escritores y especialistas en marketing utilizan ChatGPT para agilizar sus procesos creativos, optimizar flujos de trabajo y mejorar la productividad. ChatGPT sirve como asistente para generar ideas y perfeccionar trabajos iniciales, gracias a su capacidad para representar material lógico y contextualmente correcto. Esto lo convierte en un recurso esencial para los creadores de contenido. En el ámbito de la educación, ChatGPT puede generar diversos materiales didácticos, como diálogos auténticos, textos de comprensión oral y de lectura para que los profesores los utilicen en clase.

**Traducción de idiomas:** Es capaz de ayudar a traducir textos completos o específicos entre diferentes idiomas. Puede producir traducciones precisas gracias a sus capacidades de generación de idiomas. Además, es útil para traducir conversaciones informales, correos electrónicos y publicaciones en redes sociales, ya que maneja bien el lenguaje coloquial y conversacional. Los alumnos pueden utilizar ChatGPT para obtener definiciones y ejemplos del diccionario.

**Educación y tutoría:** Su capacidad para explicar conceptos de manera accesible ayuda a los estudiantes a comprender temas difíciles más allá de los materiales tradicionales. Los alumnos pueden practicar conversaciones con ChatGPT para mejorar su gramática y vocabulario, recibiendo retroalimentación útil. Además, ChatGPT tiene acceso a una amplia variedad de recursos de aprendizaje, lo que permite a los educadores recomendar materiales relevantes para enriquecer el aprendizaje de los estudiantes en diferentes áreas.

**Escritura creativa y narrativa:** Puede generar ideas para contenido dramático, desarrollar personajes y estructurar tramas. Su capacidad para producir contenido imaginativo se ha aprovechado en juegos y narrativas interactivas. Además, ChatGPT puede ayudar en la escritura de guiones, canciones, poemas e incluso novelas, ofreciendo sugerencias e ideas para el desarrollo de la trama, personajes y estructura narrativa.

Además, ChatGPT puede servir en muchas otras áreas, por ejemplo, soporte y servicio al cliente, apoyo en salud mental, escritura de código y programación, asistencia sanitaria y médica, etc.

Se puede ver que la poderosa función de generación de contenido de ChatGPT promueve su aplicación en muchos campos. Sin embargo, cabe mencionar que muchos estudios también han planteado limitaciones de ChatGPT. Tal como dicen Zhou Jie, Ke Pei, Qiu Xipeng, Huang Minlie, y Zhang Junping (2023: 4), ChatGPT todavía genera respuestas objetivamente incorrectas o sesgadas. Bowman (2022)

comparte la misma idea: “*There are still many cases where you ask ChatGPT a question, and it’ll give you a very impressive-sounding answer that’s just dead wrong.*”

En una línea similar, Kohnke Moorhouse y Zou (2023: 9) indican que ChatGPT no proporciona fuentes ni citas. Se podría argumentar que sus respuestas no son del todo originales, sino paráfrasis de fuentes que no han sido citadas adecuadamente (es decir, plagio).

Podemos aprovechar la tabla que muestran Nazir y Wang en su trabajo (2023: 7) para conocer generalmente las ventajas y desventajas de ChatGPT:

**Tabla 1. Descripción general de la estructura de los posibles pros y contras de ChatGPT**

Fuente: Las ventajas y desventajas de ChatGPT según Nazir y Wang (2023: 7)

Pros of ChatGPT	Cons of ChatGPT
·Natural Language Understanding	·Limited Context Understanding
·Accessibility	·Biased Outputs
·Versatility	·Misinformation Propagation
·Time-Saving	·Lack of Real Understanding
·Continuous Learning	·Ethical Concerns
·Language Translation	·Security Vulnerabilities
·Innovation Catalyst	·Privacy Risks
·Enhanced Learning Experiences	·Dependency on Training Data
·Creative Collaboration	·Limited Contextual Understanding
·Mental Health Support	·Over-Reliance on Training Data
·Context Retention	·Inability to Reason Causally
·Transfer Learning and Fine-tuning	·Generating Misinformation
·Massive Scale Data Processing	·Dependency on Prompt Phrasing
·Multilingual Capabilities	
·Human-like Interactions	
·Adaptability to Domains	

#### 2.1.4 ¿Qué es Gemini y qué funcionalidades tiene?

Al igual que ChatGPT, Google lanzó su propio chatbot el 7 de febrero de 2023, llamado “Google Bard AI”. Este chatbot puede interactuar de manera similar con los humanos y funcionará de forma conversacional, de la misma forma que lo hace Chat GPT. Está impulsado por el LLM *LaMDA* (Modelo de lenguaje para aplicaciones de diálogo) (Singh, Kumar y Mehra, 2023: 5). En el chatbot Bard de Google, los usuarios pueden ver una combinación de poder, inteligencia y creatividad (Ram y Verma, 2023).

El 8 de febrero de 2024, Google anunció oficialmente que cambiaría el nombre de su chatbot con inteligencia artificial, Bard, a Gemini. Además, anunció el lanzamiento de un nuevo plan de suscripción, Gemini Advanced, que permite a los usuarios acceder a su “modelo más potente”, Gemini Ultra 1.0, el cual compite directamente con el GPT-4 de su competidor OpenAI.

De acuerdo con Ahmed et al. (2023: 5), como herramienta colaborativa, Gemini tiene el potencial de mejorar la creatividad, aumentar la eficiencia y facilitar la implementación de conceptos innovadores. Con 137 mil millones de parámetros, Google Gemini captura una representación integral del mundo, lo que le permite ofrecer respuestas detalladas y matizadas. Sus capacidades incluyen la generación de texto, la traducción de idiomas, la creación de contenido creativo y la respuesta a preguntas informativas.

#### 2.1.5 ¿Qué diferencias hay entre ChatGPT y Gemini?

La aparición de Gemini ha provocado debates entre muchas personas en la comunidad tecnológica. En el artículo de Ahmed et al. (2023: 2) se plantea la siguiente pregunta: “*Can Google’s brainchild hold its own against the innovative prowess of ChatGPT?*”. Este autor afirma que GPT-4 genera respuestas más precisas y menos alucinaciones

que Gemini, como inteligencia artificial, algunas veces podrían producir contenidos falsos o plagiados o con información sin fundamento.

No obstante, Ram y Verma (2023: 260) opinan que, dado que ChatGPT solo ha sido entrenado con datos hasta mayo de 2023,<sup>1</sup> a menudo solo proporciona información hasta esa fecha con precisión, mientras que Google Gemini puede notificar a los consumidores sobre los acontecimientos más recientes. A partir de estas distintas opiniones, podemos hacer una comparación entre ChatGPT y Gemini.

Tal como dicen Ahmed et al. (2023: 2), ChatGPT y Gemini tienen enfoques diferentes. ChatGPT se centra más en la generación de texto, mientras que Gemini está diseñado para producir conversaciones que se asemejen a las humanas. Esto hace que Gemini funcione mejor que ChatGPT en interactuar con los usuarios en diálogos interactivos. Por otro lado, ChatGPT está más enfocado en comprender las consultas de los usuarios y responder mensajes de texto.

En el trabajo de Singh, Kumar y Mehra (2023: 5), se han marcado las diferencias entre Google Gemini AI y Chat GPT en términos de disponibilidad de datos, búsqueda de información, sesgo y precisión, detalle y explicación, accesibilidad y eficiencia. Según Singh, Kumar y Mehra, Gemini está diseñado para que cualquier persona lo pueda entender, incluidos los niños, mientras que las respuestas generadas por ChatGPT pueden ser bastante mecánicas y difíciles de entender. Pero ChatGPT puede proporcionar texto en una mayor variedad y cantidad de contextos en comparación con Gemini AI.

Singh (2022) demostró que Gemini nos permitirá editar las respuestas después de haberlas enviado, mientras que no podemos hacerlo con ChatGPT. Gemini producirá múltiples respuestas a las consultas que le plantee, a las que puede acceder

---

<sup>1</sup> Cabe señalar que se tratan de herramientas en constante actualización y evolución y que, posteriormente, se han publicado nuevas versiones de ChatGPT que superan algunas de estas limitaciones.

presionando “Ver borradores”. ChatGPT solo produce una única respuesta para cada consulta. En una línea similar se expresan Ahmed et al. (2023: 6):

*“Google’s approach to addressing limitations is evident in Gemini, where users are presented with multiple response options, called drafts, allowing for exploration and selection of the most resonant answer. In contrast, ChatGPT typically provides a single response by default, although it can generate various versions upon request. This distinction in response presentation can impact user satisfaction.”*

Según Naveed et al. (2023: 29), quienes han hecho una comparación de rendimiento de los LLMs en varias tareas de NLU (*Natural Language Understanding*) y NLG, en la habilidad de comprensión y generación de historias, GPT-4 lidera con una puntuación de 95.3, seguido por Gemini Ultra en segundo lugar con una puntuación de 87.8. Sin embargo, en cuanto a la resolución de problemas y razonamiento lógico, Gemini Ultra lidera con una puntuación de 74.4, mientras que GPT-4 se sitúa en segundo lugar con una puntuación de 67.

Podemos concluir que ChatGPT y Gemini presentan ventajas distintas. A partir de la revisión de la literatura, hemos elaborado una tabla para ilustrar las diferencias específicas entre ambos:

**Tabla 2. Comparación de las características de ChatGPT y Gemini**

Fuente: Creación propia basada en los artículos de Singh (2022) y Singh, Kumar y Mehra (2023) e información actualizada por parte de la autora del trabajo

Característica	ChatGPT	Gemini
Compañía	OpenAI	Google
Disponibilidad de datos <sup>2</sup>	Datos hasta octubre de 2023	Últimos datos almacenados

<sup>2</sup> Información obtenida después de preguntar directamente a ChatGPT y Gemini el día 17 de julio de 2024.

<b>Modelos de lengua</b>	GPT-3.5/GPT-4 (ChatGPT Plus)/GPT-4o	Gemini 1.0/ Gemini 1.5 Pro
<b>Parámetros del modelo (hasta julio de 2024)</b>	175 mil millones de parámetros	137 mil millones de parámetros
<b>Idioma soportado</b>	Más de 50 idiomas (incluidos chino simplificado y tradicional)	Más de 40 idiomas (incluidos chino simplificado y tradicional)
<b>Acceso a Internet en tiempo real</b>	Acceso a Internet en tiempo real sólo con Chat GPT-4 y GPT-4o	Acceso a Internet en tiempo real (posibilidad de consultar Google en tiempo real, por ejemplo)
<b>Tiempo de espera para la respuesta</b>	Puede tener un tiempo de reacción más lento	Reacciona rápidamente a su entorno
<b>Formato de respuesta</b>	Respuesta única, no editable después de enviar	Múltiples respuestas, respuestas editables
<b>Límite de palabras</b>	Aproximadamente 500 palabras o 4000 caracteres	Sin límite de palabras
<b>Generación de imágenes</b>	Chat GPT-3.5, no Chat GPT-4, sí, porque incorpora Dall·e Chat GPT-4o, sí	Sí
<b>Capacidades de codificación</b>	Sí	Sí
<b>Leer en voz alta</b>	Sí, soporta solo 37 idiomas a voz y 5 voces	Sí, soporta más de 40 idiomas a voz
<b>Coste</b>	ChatGPT-3.5 Gratis ChatGPT-4o Gratis pero con consultas limitadas ChatGPT-4 cuesta \$20 al mes	Gemini Gratis Gemini Advanced \$19.99 al mes

Ahmed et al. (2023: 10) llegan a la siguiente conclusión:

*“On one side we have Gemini, armed with the ability to tap into the vast expanse of the internet, stands ready to provide us with real-time answers to the burning questions. Meanwhile, ChatGPT emerges as the maestro of language generation, ready to compose eloquent prose or spin imaginative tales at your command. It’s like having a virtual Shakespeare at your beck and call, summoning words from thin air to weave captivating stories and engage in deep, thought-provoking conversations.”*

## **2.2 Estudios o experiencias previas en la aplicación de las herramientas de IA en enseñanza de lenguas**

Con el avance de la tecnología, la entrada de la Inteligencia Artificial en la educación ha sido una tendencia inevitable. Tal y como dice Chaka (2023: 1), aunque se encontraron pruebas limitadas de la colaboración entre las IA conversacionales y los profesores humanos, la enseñanza de idiomas en el futuro debería integrar las IA conversacionales para promover la amplificación de la inteligencia y disminuir la carga de trabajo de los profesores humanos. Hong (2023: 40) demostró que para los docentes, la calificación mediante IA puede liberarlos de la pesada carga de calificaciones, permitiéndoles dedicar más tiempo a la planificación de las lecciones.

En lo que concierne a las funciones de utilizar herramientas de IA en el aprendizaje del idioma, según Kohnke, Moorhouse y Zou (2023: 3), los estudiantes pueden usar ChatGPT para obtener definiciones y ejemplos de diccionario. También pueden emplearlo para escribir diálogos o producir diversos géneros de texto sobre un tema específico. Luego, ChatGPT puede ajustar la complejidad del diálogo para adaptarlo mejor a principiantes o avanzados. Esto permitiría a los profesores diferenciar los materiales para sus alumnos. De acuerdo con Baskara (2023: 347), en el ámbito del aprendizaje de idiomas, la habilidad de ChatGPT para generar diálogos realistas puede ofrecer a los estudiantes ejemplos auténticos de cómo se utiliza el idioma. Cabe señalar que Hong (2023: 40) también planteó algunas de las principales áreas que deberían atraer la atención de los investigadores para desarrollar más estudios, entre las cuales se encuentra los efectos de chatear con ChatGPT en la adquisición de las cuatro habilidades (hablar, escuchar, leer y escribir) por parte de los estudiantes.

Además de hablar del papel de la Inteligencia Artificial en el aprendizaje de idiomas, varios investigadores han realizado algunas prácticas en el aula para investigar el efecto del uso de IA. Li Xiaying et al. (2023) han realizado una investigación para explorar el potencial de ChatGPT en el apoyo y la capacitación de los estudiantes de

lengua china (*CLLs-Chinese language learners*) cuya primera lengua es el inglés para mejorar sus habilidades de escritura, los resultados destacan que ChatGPT se muestra prometedor como una herramienta de apoyo para *CLLs* de familias con bajos ingresos, reduciendo la desigualdad educativa y promoviendo el acceso equitativo a las oportunidades de aprendizaje de idiomas.

Por otra parte, de acuerdo con el trabajo de Huang Shanshan (2023), se pidió a 10 participantes que rellenaran un diario de aprendizaje del español durante siete días, que cotejaran las notas de las conversaciones con ChatGPT y que rellenaran un cuestionario después de su estudio. Los resultados revelan que ChatGPT puede ser una herramienta útil para ayudar al aprendizaje de idiomas, especialmente en las áreas de vocabulario, escritura y traducción.

Otro caso interesante es el estudio hecho por Salguero Romero (2023), que se basa en una serie de actividades realizadas a un grupo de alumnos en una clase de lengua extranjera de inglés mediante la ayuda de ChatGPT. Según la investigación, ChatGPT es flexible y puede imitar el tono de un niño de 6 años en el diálogo con los alumnos. Puede realizar actividades con los alumnos de forma independiente e iniciar nuevas actividades a partir de la actividad anterior. También es creativo y puede estimular el interés de sus alumnos.

Sin embargo, resulta relevante destacar que, la aplicación del uso de Inteligencia Artificial a la enseñanza de idiomas puede brindar algunos problemas. Según Borji (2023), debido a que ChatGPT es un modelo de aprendizaje no supervisado, tiene el potencial de generar información incorrecta o engañosa. Hong (2023: 53) indica que, para la revisión del profesor, el contenido generado por IA suele ser superficial y, en ocasiones, contiene hechos falsos y referencias inventadas. Por lo tanto, no debería ser fácil para un ensayo generado por IA pasar el escrutinio de un profesor que lo lee con atención. Además, tal como dicen Kohnke, Moorhouse y Zou (2023: 9) la mayoría de los textos de la base de datos proceden del corpus inglés y luego se traducen a la lengua de destino (por ejemplo, chino o japonés). Para los alumnos que no aprenden

inglés, puede haber información errónea. Además, la base de datos contiene muchas palabras que son más escritas que habladas. Estas cuestiones son especialmente preocupantes en la enseñanza de idiomas. Al mismo tiempo, los algoritmos tienen un sesgo cultural y pueden repercutir en los estudiantes de distintos orígenes culturales. Tal vez ellos no sean conscientes de que ChatGPT y otras herramientas basadas en IA no son culturalmente neutrales. Hong (2023: 38) también plantea una pregunta interesante: *“As ChatGPT draws information from existing sources without considering copyrights, is it ethical for teachers to use the lesson plans and materials produced by ChatGPT?”*.

En resumen, muchos estudios señalan el potencial prometedor de aplicar la inteligencia artificial en la enseñanza de lenguas. Sin embargo, creemos que se necesita más investigación práctica sobre cómo implementar el software de IA en la educación. En particular, cómo diseñar cursos y actividades, crear materiales didácticos e introducir orientación adecuada. También se debe estar atentos a posibles problemas que puedan surgir, con el fin de proporcionar a los docentes más recursos útiles en la enseñanza.

### **3. MARCO TEÓRICO: la comprensión oral en CLE**

Después de analizar el estado actual de la Inteligencia Artificial, nos enfocamos en entender los conceptos teóricos sobre la comprensión oral en CLE. En esta sección, examinaremos cómo se evalúa la comprensión oral según el MCER (*Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación.*) y HSK (*Hanyu Shuiping Kaoshi*), así como los niveles de habilidad auditiva requeridos para los estudiantes. Esta revisión es esencial para determinar si el contenido generado por la inteligencia artificial se adecúa a los niveles de los marcos más habituales. También revisaremos los estudios previos sobre la enseñanza de comprensión oral en CLE (Chino como lengua extranjera) en China y otros países en los que se destaca la importancia de la comprensión oral en el aula de CLE y refuerza a relevancia y la aplicabilidad práctica de esta investigación en el ámbito de la enseñanza de idiomas.

#### **3.1 La recomendación del Marco Común Europeo de Referencia para las Lenguas**

Según Cádiz Deleto (2002), el *Marco común europeo de referencia (MCER) para las lenguas: aprendizaje, enseñanza, evaluación* forma parte esencial del proyecto general de política lingüística del Consejo de Europa, que ha desarrollado un considerable y bien fundamentado esfuerzo por la unificación de directrices para el aprendizaje y la enseñanza de lenguas dentro del contexto europeo. Hemos encontrado muchos sitios en los que se habla sobre comprensión oral en el MCER.

En cuanto a la definición de comprensión oral, tal como dice el Consejo de Europa (2002: 96), “se trata de un proceso, donde el usuario o alumno recibe, de nuevo con cierta distancia, un fragmento hablado, que no precisa respuesta alguna.” En las actividades de comprensión auditiva, el usuario de la lengua como oyente recibe y procesa una información de entrada (*input*) en forma de enunciado emitida por uno o más hablantes (Consejo de Europa, 2002: 68).

La comprensión oral juega un papel importante en el aprendizaje del idioma. De acuerdo con el Consejo de Europa (2002: 14), la habilidad del alumno o usuario para

comunicarse en un idioma se activa cuando realiza diversas actividades lingüísticas, cada uno de estos tipos de actividades se lleva a cabo en relación con textos orales o escritos. La comprensión y la expresión (ya sea oral o escrita) son procesos primarios, ya que ambos son necesarios para la interacción. Igualmente, se divide la capacidad de aprender en variadas de aspectos de la competencia existencial, de los conocimientos declarativos, de las destrezas y de las habilidades. Las destrezas de comprensión auditiva forman parte clave en la competencia existencial, en particular, el hecho de prestar atención a lo que se dice, la conciencia acentuada de los riesgos que suponen los malentendidos culturales en las relaciones con los demás (Consejo de Europa, 2002: 12).

En el Capítulo 4, el MCER sugiere recomendaciones sobre las actividades de comprensión oral:

- escuchar declaraciones públicas (información, instrucciones, avisos, etc.);
- escuchar medios de comunicación (radio, televisión, grabaciones, cine);
- escuchar conferencias y presentaciones en público (teatro, reuniones públicas, conferencias, espectáculos, etc.);
- escuchar conversaciones por casualidad, etc.

Sumado a esto, se menciona qué habilidades de comprensión auditiva deben tener los estudiantes de diferentes niveles del idioma (Véase Tabla 3 para la lista completa). Por ejemplo, para los alumnos de A1, tienen que comprender discursos que sean muy lentos, que estén articulados con cuidado y con las suficientes pausas para asimilar el significado. Para un alumno de B2, los requisitos son relativamente altos: Comprende cualquier tipo de habla, tanto conversaciones cara a cara como discursos retransmitidos, sobre temas, habituales o no, de la vida personal, social, académica o profesional.

**Tabla 3. Competencia de comprensión auditiva para cada nivel en general por MCER**

Fuente: *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* (Consejo de Europa, 2002: 69)

<b>COMPRESIÓN AUDITIVA EN GENERAL</b>	
<b>C2</b>	No tiene dificultad en comprender cualquier tipo de habla, incluso a la velocidad rápida propia de los hablantes nativos, tanto en conversaciones cara a cara como en discursos retransmitidos.
<b>C1</b>	<p>Comprende lo suficiente como para seguir un discurso extenso sobre temas abstractos y complejos que sobrepasan su especialidad, aunque puede que tenga que confirmar algún que otro detalle, sobre todo, si no está acostumbrado al acento.</p> <p>Reconoce una amplia gama de expresiones idiomáticas y coloquiales, y aprecia cambios de registro.</p> <p>Es capaz de seguir un discurso extenso incluso cuando no está claramente estructurado, y cuando las relaciones son sólo supuestas y no están señaladas explícitamente.</p>
<b>B2</b>	<p>Comprende cualquier tipo de habla, tanto conversaciones cara a cara como discursos retransmitidos, sobre temas, habituales o no, de la vida personal, social, académica o profesional. Sólo inciden en su capacidad de comprensión el ruido excesivo de fondo, una estructuración inadecuada del discurso o un uso idiomático de la lengua.</p> <p>Comprende las ideas principales de un discurso complejo lingüísticamente que trate tanto temas concretos como abstractos pronunciados en un nivel de lengua estándar, incluyendo debates técnicos dentro de su especialidad.</p> <p>Comprende discursos extensos y líneas complejas de argumentación siempre que el tema sea razonablemente conocido y el desarrollo del discurso se facilite con marcadores explícitos.</p>
<b>B1</b>	<p>Comprende información concreta relativa a temas cotidianos o al trabajo e identifica tanto el mensaje general como los detalles específicos siempre que el discurso esté articulado con claridad y con un acento normal.</p> <p>Comprende las principales ideas de un discurso claro y en lengua estándar que trate temas cotidianos relativos al trabajo, la escuela, el tiempo de ocio, incluyendo breves narraciones.</p>
<b>A2</b>	<p>Comprende lo suficiente como para poder enfrentarse a necesidades concretas siempre que el discurso esté articulado con claridad y con lentitud.</p> <p>Comprende frases y expresiones relacionadas con áreas de prioridad inmediata (por ejemplo, información personal y familiar muy básica, compras, lugar de residencia, empleo) siempre que el discurso esté articulado con claridad y con lentitud.</p>

<b>A1</b>	Comprende discursos que sean muy lentos, que estén articulados con cuidado y con las suficientes pausas para asimilar el significado.
-----------	---

Asimismo, en base a los diferentes tipos de actividad mencionadas anteriormente, el MCER muestra requisitos más detallados para diferentes niveles, tal como se observa en la Tabla 4.

**Tabla 4. Competencia de comprender conversaciones entre hablantes nativos para cada nivel**

Fuente: *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación* (Consejo de Europa, 2002: 69)

<b>COMPRENDER CONVERSACIONES ENTRE HABLANTES NATIVOS</b>	
<b>C2</b>	Como C1.
<b>C1</b>	Puede seguir con facilidad conversaciones complejas entre terceras personas en debates de grupo, incluso sobre temas abstractos, complejos y desconocidos.
	Puede seguir conversaciones animadas entre hablantes nativos.
<b>B2</b>	Capta con algún esfuerzo gran parte de lo que se dice a su alrededor, pero puede resultarle difícil participar con eficacia en una discusión con varios hablantes nativos si no modifican su discurso de algún modo.
<b>B1</b>	Puede seguir generalmente las ideas principales de un debate largo que ocurre a su alrededor, siempre que el discurso esté articulado con claridad en un nivel de lengua estándar.
<b>A2</b>	Identifica generalmente el tema sobre el que se discute, siempre que se lleve a cabo con lentitud y claridad.
<b>A1</b>	No hay descriptor disponible.

Se puede observar en la tabla que a medida que el nivel de dominio del idioma aumenta, los estándares para la capacidad de comprensión oral también se elevan. Los criterios de evaluación se van perfeccionando constantemente en aspectos como el contenido temático, la pronunciación, la velocidad del discurso, la entonación, el acento y la duración de los materiales de audio.

Por otra parte, el MCER también cita otros sistemas de evaluación de idiomas. Por ejemplo, DIALANG (Consejo de Europa, 2002: 222) y ALTE (Consejo de Europa, 2002: 242) (*Association of Language Testers in Europe*), en los que la comprensión auditiva es un elemento clave para evaluar en todos los sistemas evaluables. Junto con comprensión de lectura, expresión escrita y expresión oral, se consideran las cuatro destrezas fundamentales del aprendizaje de idiomas.

### 3.2 La importancia de la comprensión oral en el HSK

*El Hanyu Shuiping Kaoshi* (HSK), es una prueba estandarizada internacional que se centra en evaluar la capacidad para utilizar el chino para comunicarse en la vida, el estudio y el trabajo de los candidatos cuyo primer idioma no es el chino (Zhongwen Kaoshi Fuwu Wang). Tal como mencionan He Wanying y Han Lu (2024: 176), la prueba del HSK goza de gran conocimiento y buena reputación entre los estudiantes de chino, y se considera una forma eficaz de evaluar el dominio del chino<sup>3</sup>. También es la prueba que exige el gobierno chino para poder solicitar becas o acceder a estudios avanzados en China y, en este sentido, se puede considerar que es un trámite necesario para demostrar de manera oficial que se tiene cierto nivel de competencia en chino.

El antiguo examen HSK se divide principalmente en 6 niveles, desde HSK Nivel 1 (el nivel más bajo) hasta HSK Nivel 6 (el nivel más alto). Los requisitos de los exámenes en los diferentes niveles varían, pero en general, los contenidos principales de la prueba incluyen 3 partes, que son *tingli* (听力) ‘Comprensión oral’, *yuedu* (阅读) ‘Comprensión lectora’ y *shuxie* (书写) ‘Exposición escrita’. Según Zhang Jinjun et al. (2010), en la primera y segunda versión del examen HSK en cada nivel, la proporción de las preguntas de comprensión oral está entre el 40% y casi el 60% de todas las

---

<sup>3</sup> Texto original traducido por la autora de este trabajo: « HSK 考试在中文学习者中具有较高的认知度和较好的口碑，而且被看作是一种有效的中文能力测试方式。 » (He Wanying y Han Lu, 2024)

preguntas.<sup>4</sup> En una línea similar se expresan He Wanying y Han Lu (2024, 176), la prueba de comprensión auditiva es la primera parte del HSK y representa un tercio de la puntuación total. Desempeña un papel importante y especial en el cultivo de las habilidades de comunicación oral de los estudiantes.<sup>5</sup> Además, en términos de la cantidad de preguntas, en todos los niveles del examen HSK anterior, las preguntas de comprensión oral superan en número a las de lectura y escritura, ocupando la mayor proporción de puntos.

A pesar del uso extendido del HSK como examen para certificar el conocimiento del chino como lengua extranjera, muchos consideraron que el nivel 6 del HSK no representaba el nivel más alto de dominio de la lengua china y solo equivalía al nivel B2-C1 del MCER. Tal como dicen Casas-Tost y Rovira-Esteva (2018: 18), la equivalencia entre los niveles del MCER y del HSK fue cuestionada por la Asociación de profesores de chino de los países de habla alemana, estableciendo una tabla de equivalencias alternativa más ajustada, donde por ejemplo el nivel HSK 3 se corresponde con el A1 del MCER. Como resultado, el nuevo examen HSK ha sido revisado para garantizar que esté más en línea con los estándares internacionales ampliamente aceptados de competencia lingüística.

El nuevo HSK constará de tres grandes etapas: principiante, intermedio y avanzado. Cada una de estas etapas se dividirá en tres niveles, con un total de nueve niveles. Los seis primeros niveles del nuevo HSK tienen sus propios requisitos de vocabulario, caracteres chinos, sílabas y número de sílabas. Los tres últimos niveles pertenecen a la etapa avanzada, que aumenta el examen para incluir también las destrezas de traducción y está dirigido a los estudiantes que quieran cursar un máster o un doctorado en China.

---

<sup>4</sup> Texto original traducido por la autora de este trabajo: «以 HSK 考试为例, 旧版和新版的各个级别的考试中, 听力的比重在 40% 到近 60% 之间。» (Zhang Jinjun et al., 2010)

<sup>5</sup> Texto original traducido por la autora de este trabajo: «HSK 听力考试是 HSK 的第一部分, 占总成绩三分之一, 旨在测试考生的听力理解能力, 在培养学生言语交际能力上具有重要且特殊的地位。» (He Wanying y Han Lu, 2024)

De acuerdo con el Xin HSK Yi Ji Kaoshi Dagang (新 HSK 一级考试大纲) ‘Nuevo plan de estudios del examen HSK Nivel 1’, los estudiantes que aprueban el HSK 1 tienen la capacidad de entender y usar algunas palabras y frases chinas muy simples para satisfacer necesidades de comunicación básicas. El examen consta de un total de 20 preguntas de comprensión auditiva, que implican escuchar frases, oraciones o conversaciones y seleccionar imágenes o respuestas a las preguntas. El nivel de dificultad es relativamente simple y básico. Por el contrario, el Xin HSK Liu Ji Kaoshi Dagang (新 HSK 六级考试大纲) ‘Nuevo plan de estudios del examen HSK Nivel 6’ indica que los candidatos que aprueban el HSK 6 tienen la capacidad de comprender fácilmente la información china que escuchan o leen, y expresar sus opiniones de forma oral o escrita con fluidez. El examen de comprensión auditiva consta de un total de 50 preguntas, donde cada pregunta se reproduce solo una vez y los estudiantes deben seleccionar las respuestas de varios pasajes y entrevistas. El nivel de dificultad es considerablemente más alto que en niveles anteriores.

Con respecto al diseño de las preguntas de comprensión oral del HSK, Liu Songhao (2001: 104) destacó en su artículo que en el HSK Básico hay un total de 50 preguntas de comprensión auditiva, de las cuales 15 emplean el método de convertir información de texto en imágenes. Esto se hace para evitar posibles problemas para los estudiantes que no pueden comprender las opciones presentadas. Por otro lado, el HSK Avanzado en comprensión auditiva agrega contenido de audios en contexto real y presta más atención a la comunicación en situaciones naturales.<sup>6</sup> Al mismo tiempo, con el propósito de evaluar las habilidades auditivas reales de los alumnos, la prueba HSK utiliza materiales auditivos con temas diversos, géneros variados y diferentes duraciones (He Wanying y Han Lu, 2024: 177).

---

<sup>6</sup> Texto original traducido por la autora de este trabajo: «基础 HSK 共有 50 个听力试题，其中 15 个采用将文字信息转换为图像的方法，避免了看不懂选项对学生可能产生的影响。高等 HSK 听力理解增加了实况录音的内容，更加重视自然状态下的交际。」 (Liu Songhao, 2001: 104)

Estas características demuestran la ingeniosa planificación de las preguntas de comprensión oral en HSK, lo que resalta la relevancia de la habilidad auditiva en dicho examen.

### **3.3 Estudios previos sobre la enseñanza de comprensión oral en CLE**

Al revisar la literatura relevante sobre la comprensión oral en la prueba HSK y en la enseñanza de chino como lengua extranjera, encontramos que, en comparación con la lectura y la escritura, la comprensión auditiva es la parte más desafiante para los estudiantes chinos.

Según Shen Yan (1998: 143-154), a través del análisis de los resultados del examen HSK de estudiantes de intercambio de Japón y Corea del Sur en la Universidad de Tsinghua, se observó que la habilidad de comprensión auditiva de los estudiantes de ambos países era inferior a su capacidad de lectura. Específicamente, los estudiantes japoneses mostraron un rendimiento notablemente más bajo en la sección de comprensión oral, con un 89% de ellos obteniendo puntajes más bajos en comparación con la comprensión de lectura. Duan Junyu (2022: 93) comparte la misma opinión:

*“Through the two years of professional Chinese study in their freshman and sophomore years, students can basically reach HSK-3 level in Chinese, and most of them have good reading and writing skills in Chinese, but their listening and speaking skills are a bit inferior.”*

Por consiguiente, la mejora de la capacidad de comprensión oral representa un desafío significativo para los estudiantes de chino y ha suscitado discusiones por investigadores en el campo de CLE.

Cao Hui (2000: 11) concluye que los estudios de la comprensión oral en CLE tratan sobre diferentes temas, como los libros usados, las actividades, la pedagogía, las dificultades para entender oralmente, las habilidades de escucha y la rapidez del habla.

Durante nuestro proceso de revisión bibliográfica, confirmamos este punto. Al integrar estudios previos y actuales sobre la enseñanza del chino como lengua extranjera tanto en China como en otros países, en los próximos párrafos abordaremos principalmente cuatro áreas: análisis de las dificultades en la comprensión oral para los estudiantes extranjeros, redacción del material, estrategias de enseñanza en el aula y aplicación de nuevas tecnologías para mejorar la enseñanza de la comprensión oral en CLE.

De acuerdo con Du Shisheng (2013), el bajo interés de los estudiantes, la débil capacidad de pronunciación, el vocabulario insuficiente, la falta de conocimiento gramatical y cultural son las razones de los problemas de comprensión oral. Tal como explica Shen Yan (1998: 143), a los alumnos les resulta difícil adaptarse al examen HSK debido a: (1) la falta de formación en comprensión y expresión oral; (2) la falta de adaptación al estilo lingüístico, que es diferente del chino formal utilizado en el aula y en la vida cotidiana; (3) la falta de formación en comprensión oral con materiales específicos para el HSK. Según Yu Na (2007: 12), entre los diversos factores que influyen en la comprensión oral de los alumnos figuran el vocabulario, el habla, la entonación, la velocidad del habla, la gramática, el material, la emoción y la cultura, y destacan la importancia del vocabulario. De igual modo, Dong Xiaodai y Xie Wei (2012: 3) mencionan que las palabras nuevas son dificultad reconocida por todos los alumnos en la comprensión oral.

En relación con la redacción del material, Shi Weidong (2000) afirma:

“Hay que acortar la longitud de los textos de comprensión oral y aumentar el número de páginas y la variedad de temas. El problema del vocabulario también

afecta a la elección del corpus. El problema del vocabulario consta de dos aspectos: el ámbito y la categoría del vocabulario y la cantidad de vocabulario.”<sup>7</sup>

Acerca de las estrategias de enseñanza en el aula, Ko (2016: 98) sostiene que la mayoría de los métodos de enseñanza de la comprensión oral en chino tienden a favorecer los ejercicios mecanizados, la práctica fuera de contexto y a ignorar el carácter práctico de las situaciones comunicativas. Tal como señala Jin Lixin (1998), las clases de comprensión oral deben hacer hincapié en la racionalidad y en la importancia de utilizar las leyes de la fonética para guiar a los alumnos, en lugar de limitarse a repetir escuchas muertas. Coincidimos con Ma Xiuli, Gong Yang, Gao Xuesong y Xiang Yiqing (2017: 14) en que, los profesores de CLE deberían animar a los alumnos a tomar la iniciativa y proporcionarles más tareas de autoevaluación cuando enseñen a escuchar. Además, Wang Chunyu (2022: 181-205) hace una comparación entre el uso de métodos audiovisuales y de sólo audio en el aula de CLE, y concluye que resulta mejor el uso de métodos audiovisuales para mejorar la competencia auditiva de los alumnos.

Mientras tanto, cabe mencionar que en los últimos años se han realizado algunas investigaciones sobre la integración de la tecnología en la enseñanza de la comprensión oral en CLE o ILE (Inglés como Lengua Extranjera). Por ejemplo, Lan, Yu-Ju y Liao, Chia-Ying (2018) han estudiado los efectos de la inmersión 3D en la comprensión oral de los estudiantes de CLE. El objetivo del estudio era mejorar la comprensión auditiva de los estudiantes de chino como segunda lengua utilizando contextos auténticos en *Second Life*. Veintisiete estudiantes de CLE de 4 países participaron en el estudio de 6 semanas de duración. Los resultados del análisis muestran que la comprensión oral de los estudiantes de CLE se benefició de la

---

<sup>7</sup> Texto original traducido por la autora de este trabajo: «应当缩短听力课文的长度, 增加篇幅数量和题材的多样性。生词问题也影响对语料的选择。生词问题包括两个方面: 一是生词的范围和类别, 二是生词量。» (Shi Weidong, 2000)

inmersión en 3D. Los participantes también expresaron su preferencia por recibir clases de comprensión oral en contextos de inmersión 3D. Tai y Chen (2021) han investigado el impacto de la realidad virtual móvil en la comprensión oral de los estudiantes de ILE. Los resultados sugieren que la realidad virtual móvil puede ser una herramienta útil para promover la comprensión oral en ILE y subrayan la necesidad de realizar más investigaciones sobre esta tecnología emergente para el aprendizaje de idiomas.

En general, revisando las referencias relacionadas con la enseñanza de la comprensión oral en chino como lengua extranjera, tanto en China como en Europa, se han realizado pocos estudios en los últimos años. Sin embargo, la comprensión oral, como punto difícil en la enseñanza del chino, sigue siendo digna de estudio.

## 4. METODOLOGÍA

Utilizaremos un enfoque cualitativo y nos basaremos en los ejercicios de Zhongwen Shuiping Kaoshi HSK Si Ji Yangjuan (中文水平考试 HSK 四级样卷) ‘Documento de muestra para la prueba de dominio del chino HSK nivel 4’ como referencia. A través de comandos de entrada (*prompts*), les pediremos a ChatGPT y Gemini que generen material similar al de comprensión oral del Nuevo HSK Nivel 4 y diseñen ejercicios relacionados. Luego, crearemos una rúbrica para evaluar la calidad del texto que conformará el audio para la comprensión oral, los ejercicios y la salida de voz, con diferentes niveles de evaluación. Finalmente, compararemos los resultados de ChatGPT y Gemini en estos tres aspectos para determinar cuál de las dos herramientas puede ayudar mejor a los profesores en la preparación de material de comprensión oral.

### 4.1 Recogida de datos

Dado que no se encontraron materiales de comprensión oral generados por ChatGPT y Gemini en estudios o publicaciones previas, decidimos recopilar datos primarios totalmente nuevos mediante nuestro experimento. En primer lugar, encontramos los ejercicios de la prueba de comprensión oral en Xin HSK Si Ji Kaoshi Dagang (新 HSK 四级考试大纲) ‘Nuevo plan de estudios del examen HSK Nivel 4’ (2009), que se dividen en tres partes:

**Parte 1:** Hay 10 preguntas. Cada pregunta se escucha una vez. En cada pregunta, una persona comienza con un pasaje corto y otra persona dice una frase basada en el pasaje, que también se proporciona escrita en el papel de la prueba y se pide al candidato que juzgue si es correcta o incorrecta.

**Parte 2:** Hay 15 preguntas. Cada pregunta se escucha una vez. Cada pregunta consiste en un diálogo de dos frases entre dos personas, una tercera persona hace una pregunta basada en el diálogo y se proporcionan cuatro opciones de respuesta

en el papel de examen, el candidato selecciona la respuesta basándose en lo que escucha.

**Parte 3:** Hay 20 preguntas. Cada pregunta se escucha una vez. Las preguntas de esta parte son de 4 a 5 frases de diálogo o un pasaje corto, se hacen una o dos preguntas basadas en el diálogo o pasaje, y se proporcionan 4 opciones en el papel de examen para cada pregunta, los candidatos eligen la respuesta basándose en lo que oyen.

Al mismo tiempo, el Centre for Language Education and Cooperation también ha elaborado un modelo de examen para el nivel 4 del nuevo HSK<sup>8</sup>, del que hemos extraído ejemplos de preguntas de las tres secciones de comprensión oral.

**例题 1 (判断对错) :**

听力文本:

现在我很少看电视, 其中一个原因是, 广告太多了, 不管什么时间, 也不管什么节目, 只要你打开电视, 总能看到那么多的广告, 浪费我的时间。

题目:

他喜欢看电视广告。

答案: ×

**Ejemplo 1 (Verdadero o falso):**

Texto:

Ahora rara vez veo televisión. Una de las razones es que hay demasiados anuncios, no importa la hora o el programa, siempre que enciendes el televisor, siempre verás muchos anuncios, lo cual es un desperdicio de mi tiempo.

Pregunta:

Le gusta ver anuncios de televisión.

Respuesta: ×

**例题 2 (单选题) :**

听力文本:

女: 该加油了, 去机场的路上有加油站吗?

男: 有, 你放心吧。

题目:

---

<sup>8</sup> Jiaoyubu Zhongwai Yuyan Jiaoliu Hezuo Zhongxin (教育部中外语言交流合作中心). Zhongwen Shuiping Kaoshi HSK Si Ji Yangjuan (中文水平考试 HSK 四级样卷) [Documento de muestra para la prueba de dominio del chino HSK nivel 4].

<[https://download.chinesetest.cn/newhsk-site/model/HSK4\\_yj.pdf](https://download.chinesetest.cn/newhsk-site/model/HSK4_yj.pdf)> [Fecha de consulta: 27-5-2024].

男的主要是什么意思？

选项：

A 去机场 B 快到了 C 油是满的 D 有加油站

答案：D

**Ejemplo 2 (Elección múltiple):**

Texto:

Mujer: Es hora de repostar, ¿hay alguna gasolinera de camino al aeropuerto?

Hombre: Sí, no te preocupes.

Pregunta:

¿Qué significa principalmente lo que dice el hombre?

Opciones:

A) Que van al aeropuerto B) Que están a punto de llegar C) Que el depósito está lleno D) Que hay una gasolinera

Respuesta: D

**例题 3 (单选题) :**

听力文本：

男：把这个材料复印五份，一会儿拿到会议室发给大家。

女：好的。会议是下午三点吗？

男：改了。三点半，推迟了半个小时。

女：好，602 会议室没变吧？

男：对，没变。

题目：

会议几点开始？

选项：A 14: 00 B 15: 00 C 15: 30 D 18: 00

答案：C

**Ejemplo 3 (Elección múltiple):**

Texto:

Hombre: Haz cinco copias de este material y llévalo luego a la sala de conferencias para repartirlo entre todos.

Mujer: De acuerdo. ¿La reunión es a las tres de la tarde?

Hombre: Ha cambiado. Se ha pospuesto media hora.

Mujer: Vale. La sala de conferencias 602 no ha cambiado, ¿verdad?

Hombre: Sí, no ha cambiado.

Pregunta:

¿A qué hora empieza la reunión?

Opciones:

A) A las dos B) A las tres C) A las tres y media D) A las seis

Respuesta: C

Vamos a instruir a ChatGPT y Gemini para que imiten los ejemplos de los ejercicios de comprensión oral del nuevo HSK nivel 4, generen oraciones simples, diálogos cortos y largos que contengan el vocabulario del HSK nivel 4 que queremos evaluar.

Además, ChatGPT y Gemini tendrán que diseñar ejercicios y respuestas correspondientes.

Sin embargo, es difícil introducir los comandos más eficaces para que ambos programas generen con éxito contenidos que cumplan nuestros requisitos. Tal como dice Morales-Chan (2023: 2) sobre la importancia de los *prompts*, la calidad de los *prompts* es uno de los factores más importantes para lograr una conversación exitosa en ChatGPT. A la hora de interactuar con ChatGPT, es esencial proporcionar un contexto claro y preciso. Por lo tanto, tenemos que probar muchas veces para lograr los *prompts* correctos antes de la evaluación y la comparación, y la capacidad de ChatGPT y Gemini para comprender y responder correctamente a las instrucciones también será evaluado.

El artículo de Morales-Chan (2023: 3) nos proporciona diez recomendaciones iniciales para interactuar eficazmente con ChatGPT:

- Definir un objetivo claro al momento de interactuar con Chat GPT.
- Mantener el mensaje conciso.
- Utilizar un lenguaje natural y evitar la ambigüedad en el mensaje.
- Proporcionar contexto suficiente para que Chat GPT comprenda el contexto de la conversación y pueda responder en consecuencia.
- Evitar las preguntas abiertas y utilizar palabras y frases específicas para guiar el resultado deseado.
- Tener en cuenta el tono y el lenguaje utilizado en la conversación.
- En el caso de utilizar jerga o términos técnicos, se recomienda proporcionar definiciones claras para evitar malas interpretaciones.
- Revisar cuidadosamente el mensaje antes de enviarlo, con el fin de evitar ambigüedades y garantizar que el mensaje se transmita adecuadamente a ChatGPT.

También incorporaremos dichas sugerencias en el experimento que haremos para encontrar *prompts* relativamente concisos y eficaces. Eventualmente guardaremos los

diálogos con ambos programas y conservaremos estos datos en forma de capturas de pantalla para poder analizarlos y comparar, de este modo, cuáles han sido los *prompts* más efectivos.

En total, se utilizarán 3 *prompts*, que se replicarán tanto en ChatGPT 3.5 como en Gemini, lo que nos permitirá analizar 15 ejercicios para cada herramienta, de los cuales del 1 al 5 corresponderán al primer *prompt*, del 6 al 10 corresponderán al segundo *prompt* y del 11 al 15 corresponderán al tercer *prompt*. Las fechas de recogida de datos son del 22 de mayo de 2024 al 25 de junio de 2024.

## **4.2 Análisis de los datos**

En última instancia, los datos que analizaremos serán de tres tipos. En primer lugar, el contenido textual de los materiales de comprensión oral generados por ChatGPT y Gemini. En segundo lugar, los ejercicios diseñados. Y, por último, su audio basado en las funciones de conversión de texto a voz de ChatGPT y Gemini. Diseñaremos rúbricas como herramienta de evaluación.

Será una rúbrica de tipo analítico. Según Gatica-Lara y Uribarren-Berrueta en su artículo de «Cómo elaborar una rúbrica» (2012: 62), se recomienda utilizar la rúbrica analítica cuando hay que identificar los puntos fuertes y débiles, tener información detallada, y valorar habilidades complejas. También utilizaremos el modelo de rúbrica que Gatica-Lara y Uribarren-Berrueta (2012) proporciona para diseñar nuestras rúbricas de evaluación. (Véase Tabla 5)

### **Tabla 5. Elementos de la rúbrica analítica**

Fuente: «Cómo elaborar una rúbrica» (Gatica-Lara y Uribarren-Berrueta, 2012: 62)

Conceptos/rubros	Escalas/niveles ejecución (cuantitativo/cualitativo/mixto)			
	4	3	2	1
Aspectos a evaluar	Criterios o evidencias a alcanzar	Criterios o evidencias a alcanzar	Criterios o evidencias a alcanzar	Criterios o evidencias a alcanzar

En cuanto a los pasos para la elaboración de las rúbricas, Gatica-Lara y Uribarren-Berrueta (2013: 64) indican los siguientes:

1. Determinar los objetivos del aprendizaje.
2. Identificar los elementos o aspectos a valorar.
3. Definir descriptores, escalas de calificación y criterios (Véase Tabla 5).
4. Determinar el peso de cada criterio.
5. Revisar la rúbrica diseñada y reflexionar sobre su impacto educativo.

Tal como indica, en primer lugar tenemos que decidir el objetivo de las rúbricas. Se trata de identificar los puntos fuertes y débiles de los contenidos generados por ChatGPT y Gemini, hacer una comparación entre ambos, y valorar cuál de los dos reduce trabajo al docente cuando preparemos materiales de comprensión oral en el aula de CLE para los alumnos que están preparando el examen del nuevo HSK nivel 4.

Basadas en las 5 etapas señaladas por Gatica-Lara y Uribarren-Berrueta (2013: 64), a continuación, diseñaremos las rúbricas para evaluar los textos, ejercicios y audios producidos por ChatGPT y Gemini.

#### **4.2.1 Rúbrica para evaluar los textos de la comprensión oral generados por ChatGPT y Gemini**

En cuanto a los aspectos para evaluar sobre los textos de la comprensión oral generados, dado que no encontramos en el mercado requisitos del HSK para los

materiales de comprensión oral, nos remitimos a otras bibliografías y estudios relacionados con la comprensión oral.

Primero, nos remitimos a los materiales didácticos del profesor Antonio Paoliello (2023), en el apartado sobre cómo seleccionar el texto del material de comprensión oral, Paoliello sugiere que se deba prestar atención a la tipología textual, la temática, si es genuino o auténtico. Paoliello también destaca que la dificultad del texto será determinada, generalmente, por el nivel que queremos certificar.

Según Paoliello (2023), los aspectos a tener en cuenta en la selección de textos para comprensión oral son:

- La tipología textual (monólogos, diálogos, anuncios, etc.)
- La estructura textual (expositivo, descriptivo, instructivo...)
- La temática (de ámbito personal, público, profesional, educativo)
- Si es genuino o auténtico
- La velocidad del habla
- El acento
- La extensión
- En las interacciones mejor voces femeninas y masculinas

Además, Villalba (2009: 25) enumera unas especificaciones para los diseñadores de las pruebas de ELE (español como lengua extranjera), como la naturaleza del lenguaje oral, el nivel de formalidad, el vocabulario o la gramática. Consideramos que son especificaciones extrapolables a otros exámenes de comprensión oral como HSK.

Por último, el Consejo de Europa (2002: 144-151) también describe el papel de los textos en el aprendizaje y la enseñanza de lenguas. Combinando estos elementos y puntos, creamos la siguiente rúbrica de evaluación. (Véase Tabla 6)

**Tabla 6. Rúbrica para evaluar los textos de la comprensión oral generados por ChatGPT y Gemini**

Fuente: Creación propia

<b>Rúbrica para evaluar los textos de la comprensión oral generados por ChatGPT y Gemini</b>					
<b>Criterios</b>	<b>Nivel</b>				<b>Resultado final</b>
	<b>4. Excelente</b>	<b>3. Bueno</b>	<b>2. Aceptable</b>	<b>1. Insuficiente</b>	
<b>Tipo de texto</b>	Genera con precisión el formato requerido, como frases simples o diálogos. Comprende completamente las instrucciones.	Genera con precisión el formato requerido, como frases simples o diálogos. Pero no comprende todas las instrucciones, por ejemplo, no incluye las palabras proporcionadas o está mal imitando.	Genera con precisión el Tipo de texto, pero tiene problema en comprender la mayoría de las instrucciones.	No genera el formato que cumple con los requisitos, como frases simples o diálogos.	
<b>Longitud del texto</b>	Todas las oraciones del texto tienen la longitud requerida según el esquema, expresando completamente el contenido.	La mayoría de las oraciones del texto tienen la longitud requerida según el esquema, expresando completamente el contenido.	Algunas de las oraciones del texto tienen la longitud requerida según el esquema, expresando parcialmente el contenido.	La longitud de las oraciones del texto es excesiva o insuficiente, no cumpliendo con el esquema y no expresando completamente el contenido.	

### Rúbrica para evaluar los textos de la comprensión oral generados por ChatGPT y Gemini

Criterios	Nivel				Resultado final
	4. Excelente	3. Bueno	2. Aceptable	1. Insuficiente	
<b>Adecuación al nivel indicado</b>	Todos los términos, la gramática y la dificultad de las oraciones se adecuan al nivel nuevo HSK 4.	La mayoría de los términos, la gramática y la dificultad de las oraciones se adecuan al nivel nuevo HSK 4.	Algunos términos, la gramática y la dificultad de las oraciones se adecuan al nivel nuevo HSK 4.	Los términos, la gramática y la dificultad de las oraciones son demasiado simples o demasiado complejos, no cumpliendo con el nivel nuevo HSK 4.	
<b>Corrección a nivel gramatical y ortográfico</b>	Todas las oraciones son gramaticalmente correctas y se expresan de manera clara, sin errores ortográficos.	La mayoría de las oraciones son gramaticalmente correctas y se expresan de manera clara, sin errores ortográficos.	Hay bastantes oraciones con errores gramaticales que dificultan la comprensión de las mismas.	La cantidad de errores gramaticales es muy alta, hasta el punto de que no se puede entender bien la intención comunicativa.	
<b>Naturalidad de la lengua y coherencia lógica del mensaje</b>	El lenguaje es fluido y natural, tiene lógica y se ajusta a las convenciones del lenguaje oral.	El lenguaje es en su mayoría natural, tiene lógica y se ajusta a las convenciones del lenguaje oral.	El lenguaje es algo forzado, mecánico, pero con lógica básica y se acerca a las convenciones del lenguaje oral.	El lenguaje no es natural, es muy forzado y la lógica es confusa.	
<b>Autenticidad del texto</b>	El texto es auténtico, se ajusta a los temas comunes del examen HSK y refleja la vida real.	La mayoría del texto es auténtico, se ajusta a los temas comunes del examen HSK y refleja la vida real.	Algunas partes del texto son auténticas, se ajustan a los temas comunes del examen HSK y reflejan la vida real.	El texto no se ajusta a la vida real y no refleja los temas comunes del examen HSK.	

Rúbrica para evaluar los textos de la comprensión oral generados por ChatGPT y Gemini					
Criterios	Nivel				Resultado final
	4. Excelente	3. Bueno	2. Aceptable	1. Insuficiente	
Conclusión					

#### 4.2.2 Rúbrica para evaluar los ejercicios de la comprensión oral generados por ChatGPT y Gemini

En cuanto al diseño de los ejercicios, también hemos tomado referencia de los materiales de clase del profesor Paoliello y la profesora Liao Shu-ching. Tal como dice el profesor Paoliello, en la redacción de ítems de la comprensión oral, las preguntas deben contener el lenguaje de nivel igual o inferior al texto, hay que asegurarse que las preguntas no pueden contestarse ni por conocimiento del mundo (sentido común) ni por coincidencia de léxico, además, evitar adverbios de grado que puedan tener valor subjetivo, por ejemplo, *feichang* (非常) [muy], *bijiao* (比较) [relativamente], *xiangdang* (相当) [mucho], *bu tai* (不太) [no muy]. Según profesora Liao, al redactar material auditivo, es importante evitar el exceso de vocabulario desconocido y un diseño inadecuado de las preguntas. Integrando los materiales de los dos profesores y nuestras propias ideas, hemos creado la siguiente tabla para evaluar los ejercicios creados por ChatGPT y Gemini. (Véase Tabla 7)

**Tabla 7. Rúbrica para evaluar los ejercicios de la comprensión oral generados por ChatGPT y Gemini**

Fuente: Creación propia

<b>Rúbrica para evaluar los ejercicios de la comprensión oral generados por ChatGPT y Gemini</b>					
<b>Criterios</b>	<b>Nivel</b>				<b>Resultado final</b>
	<b>4. Excelente</b>	<b>3. Bueno</b>	<b>2. Aceptable</b>	<b>1. Insuficiente</b>	
<b>Tipo de ejercicio</b>	Genera correctamente el tipo de pregunta requerido según las instrucciones, como selección múltiple o verdadero/falso. La forma en que se presenta el contenido es clara y concisa.	Genera correctamente el tipo de pregunta requerido según las instrucciones, pero la forma en que se presenta el contenido no es muy clara o concisa.	Genera correctamente el tipo de pregunta requerido según las instrucciones, pero el formato está mal presentado.	No genera el tipo de pregunta requerido.	
<b>Relación al texto</b>	Las preguntas están estrechamente relacionadas con el texto, bien diseñadas y contribuyen a evaluar la comprensión del contenido.	La mayoría de las preguntas (aunque no todas) están relacionadas con el texto, están diseñadas adecuadamente y contribuyen a evaluar la comprensión del contenido.	Algunas preguntas están relacionadas con el texto, pero el diseño y la contribución a la evaluación de la comprensión del contenido son inconsistentes.	Las preguntas no están relacionadas con el texto o no contribuyen a la evaluación de la comprensión del contenido.	
<b>Adecuación al nivel indicado</b>	Las preguntas son fáciles de entender y no contienen palabras nuevas o estructuras gramaticales más complejas que el	Las preguntas son generalmente comprensibles y solo ocasionalmente contienen	Algunas preguntas son difíciles de entender o contienen palabras nuevas o	La mayoría de las preguntas son difíciles de entender o contienen palabras nuevas o	

## Rúbrica para evaluar los ejercicios de la comprensión oral generados por ChatGPT y Gemini

Criterios	Nivel				Resultado final
	4. Excelente	3. Bueno	2. Aceptable	1. Insuficiente	
	nivel de lenguaje del texto.	palabras nuevas o estructuras gramaticales más complejas que el nivel de lenguaje del texto.	estructuras gramaticales más complejas que el nivel de lenguaje del texto.	estructuras gramaticales más complejas que el nivel de lenguaje del texto.	
<b>Diseño de las preguntas</b>	Las preguntas están bien diseñadas, no contienen términos subjetivos o respuestas ambiguas. No se puede contestarse por sentido común o por coincidencia de léxico.	La mayoría de las preguntas están bien diseñadas, con pocos términos subjetivos o respuestas ambiguas. No se puede contestarse por sentido común o por coincidencia de léxico.	Algunas preguntas tienen problemas de diseño, contienen términos subjetivos o respuestas ambiguas. Se puede contestarse por sentido común o por coincidencia de léxico.	La mayoría de las preguntas tienen problemas de diseño, contienen términos subjetivos o respuestas ambiguas. Se puede contestarse por sentido común o por coincidencia de léxico.	
<b>Diseño de las opciones y respuestas</b>	Las opciones de las preguntas de selección múltiple se encuentran en el texto o son distractores relacionados con la opción correcta, tienen una estructura y longitud consistentes y no tienen problemas gramaticales. No hay patrón ni	La mayoría de las opciones de las preguntas de selección múltiple se encuentran en el texto o son distractores relacionados con la opción correcta, tienen una estructura y longitud consistentes y no tienen	Algunas opciones de las preguntas de selección múltiple se encuentran en el texto o son distractores relacionados con la opción correcta, pero tienen estructuras o longitudes inconsistentes	La mayoría de las opciones de las preguntas de selección múltiple no se encuentran en el texto, no se establecen opciones de distractores, tienen estructuras o longitudes inconsistentes y problemas	

<b>Rúbrica para evaluar los ejercicios de la comprensión oral generados por ChatGPT y Gemini</b>					
<b>Criterios</b>	<b>Nivel</b>				<b>Resultado final</b>
	<b>4. Excelente</b>	<b>3. Bueno</b>	<b>2. Aceptable</b>	<b>1. Insuficiente</b>	
	repetición en las opciones o respuestas.	problemas gramaticales. No hay patrón ni repetición en las opciones o respuestas.	o problemas gramaticales. Hay patrón o repetición en las opciones o respuestas.	gramaticales. Hay patrón o repetición en las opciones o respuestas.	
<b>Conclusión</b>					

#### **4.2.3 Rúbrica para evaluar la dicción oral de la comprensión oral generados por ChatGPT y Gemini**

En esta parte, nos referimos al artículo de Zanón (2010) que se llama «Claves para comprender la destreza de la comprensión oral en lengua extranjera» y los materiales didácticos del profesor Paoliello (2023). Zanón (2010: 197) opina que hay unas variables relativas al hablante que pueden facilitar o dificultar la comprensión por parte del oyente: la velocidad, el acento, la voz del hablante, el sexo, etc. Paoliello también destaca la importancia de evitar la voz monótona. Además, también hemos tenido en cuenta posibles problemas al generar material de voz con inteligencia artificial, como la mecanización, la falta de precisión en la pronunciación y la falta de naturalidad. Por otro lado, consideramos la conveniencia para los profesores al reproducir el audio. Estos aspectos también se han integrado en nuestra rúbrica. (Véase Tabla 8)

Tenemos en cuenta que el sonido es una función relativamente estable y no produce tantas posibilidades como el texto, así que hicimos algunos cambios en las reglas y las modificamos para que se calificaran según el cumplimiento de las reglas.

Para cada elemento de evaluación, se pueden utilizar los siguientes cuatro niveles de puntuación:

**1.Insuficiente:** No cumple con los requisitos básicos y el rendimiento es muy insuficiente.

**2.Aceptable:** Básicamente cumple con los requisitos, pero aún existen algunos problemas o deficiencias.

**3.Bueno:** Se alcanza un alto nivel, tiene un buen desempeño y cumple con las expectativas.

**4.Excelente:** Excelente desempeño, cumple plenamente o supera las expectativas.

**Tabla 8. Rúbrica para evaluar la dicción oral de la comprensión oral generados por ChatGPT y Gemini**

Fuente: Creación propia.

<b>Rúbrica para evaluar la dicción oral de la comprensión oral generados por ChatGPT y Gemini</b>		
<b>Criterios</b>	<b>Descripción</b>	<b>Resultado final</b>
<b>Integridad del contenido</b>	Lee el contenido original palabra por palabra, sin omitir ningún carácter ni añadir caracteres inexistentes en la versión escrita.	
<b>Precisión de la pronunciación</b>	La pronunciación es precisa y se ajusta al del mandarín estándar, incluyendo una correcta articulación de las consonantes iniciales y finales, las vocales y los tonos.	
<b>Velocidad de habla</b>	El habla tiene un ritmo moderado y puede ajustarse según el contenido del texto, sin ser demasiado rápido ni demasiado lento, manteniendo un flujo de habla fluido.	
<b>Cambio de tonos especiales</b>	Puede manejar correctamente tonos especiales, como cuando “不” está seguido por el cuarto tono, se pronuncia “bú” y no “bù”. Cuando “一” está seguido por el cuarto tono, se pronuncia “yī” en lugar de “yǐ”. Pronuncia correctamente las palabras con tono neutro, como la segunda sílaba de “baba 爸爸 [padre]” y “mama 妈妈[madre]”.	
<b>Entonación</b>	Puede transmitir con precisión las emociones y el tono del texto original, como preguntas, sorpresa,	

	preocupación, etc., para que los oyentes puedan sentir el matiz emocional del texto.	
<b>Naturalidad de habla</b>	La fluidez de habla es natural, con una conexión suave entre las oraciones, sin pausas o interrupciones evidentes, lo que hace que los oyentes sientan que están interactuando con una persona real.	
<b>Calidad de la voz</b>	El sonido es claro y melodioso, sin ruidos ni interrupciones, y ofrece opciones de voz masculina y femenina para satisfacer las necesidades de diferentes usuarios.	
<b>Facilidad de la operación</b>	Amigable para el uso de los profesores, proporciona botones de operación convenientes, como la capacidad para pausar o reanudar la reproducción del sonido en cualquier momento.	
<b>Conclusión</b>		

## **5. ANÁLISIS DE LOS MATERIALES**

De acuerdo con el marco teórico y la metodología establecidos, en esta sección vamos a proporcionar las mismas instrucciones tanto a ChatGPT 3.5 como a Gemini. Las instrucciones exigirán que imiten el tipo de ejercicios que se encuentran en el documento de muestra para la prueba de dominio del chino HSK nivel 4 (Ejemplos 1, 2 y 3 mencionados en la parte de recogida de datos). Revisaremos los resultados generados por ambos, y evaluaremos el contenido utilizando las rúbricas diseñadas específicamente para este propósito. Esta parte se enfocará principalmente en el análisis de los materiales generados por ChatGPT y Gemini, incluyendo textos, ejercicios y la dicción oral.

Primero, presentaremos el contenido generado por ChatGPT y procederemos a evaluarlo. A continuación, haremos lo mismo con el contenido generado por Gemini. Finalmente, compararemos los resultados de las evaluaciones de ambos mediante las rúbricas, con el objetivo de determinar cuál software de inteligencia artificial presenta un mejor desempeño en la generación de materiales de comprensión oral de CLE.

### **5.1 Análisis de ChatGPT**

Durante el proceso de la recogida de datos con ChatGPT, optimizamos continuamente las instrucciones proporcionadas, lo que nos permitió generar un contenido que consideramos adecuado para una evaluación. Las instrucciones finales incluyeron lo siguiente: mostrar un ejemplo de examen del nuevo HSK4, pedir a ChatGPT que genere cinco preguntas de comprensión oral similares, con un límite de palabras según los requisitos del esquema o conclusiones por Liang Wenjiao (2015: 8-9), y que incluya vocabulario del nuevo HSK4. Finalmente, se especificó si las preguntas eran de opción múltiple o de verdadero/falso.

A continuación, se presentará los materiales de comprensión oral y las preguntas generadas por ChatGPT utilizando estas instrucciones.

### 5.1.1 Análisis de los textos y ejercicios de verdadero y falso generado por ChatGPT 3.5 (Ejemplo 1)

Primero, mostramos los resultados de generación de las preguntas de verdadero y falso en la primera parte de comprensión oral del examen del nuevo HSK Nivel 4. Dado el requisito de que el contenido generado por ChatGPT esté en chino, también utilizamos la entrada de comandos en chino para garantizar la uniformidad lingüística.

#### Extracto 1. Ejercicios de verdadero y falso generado por ChatGPT 3.5

##### PROMPT:

听力材料：现在我很少看电视，其中一个原因是，广告太多了，不管什么时间，也不管什么节目，只要你打开电视，总能看到那么多的广告，浪费我的时间。题目：他喜欢看电视广告。答案：×。这是一道新 HSK 四级考试的听力材料和题目，请你模仿它，生成 5 道听力题，听力材料 30 到 60 字，材料中分别包含以下词汇：爱情、安排、安全、按时、按照。题目答案有对有错。

Material para escuchar: ahora rara vez veo televisión. Una de las razones es que hay demasiados anuncios, no importa la hora o el programa, siempre que enciendas el televisor, siempre verás muchos anuncios, lo cual es un desperdicio de mi tiempo. Pregunta: Le gusta ver anuncios de televisión. Respuesta: ×. Este es un material de escucha y una pregunta para el nuevo examen HSK Nivel 4. Imítalo y genera 5 preguntas de comprensión oral. El material tiene de 30 a 60 caracteres. El material contiene las siguientes palabras: *aiqing* (爱情) [amor], *anpai* (安排) [arreglar], *anquan* (安全) [seguridad], *anshi* (按时) [puntualmente] y *anzhao* (按照) [de acuerdo con]. Hay respuestas correctas e incorrectas a las preguntas.

##### ChatGPT:

听力材料 1：我认为爱情是生活中美好的一部分，但是对我来说，最重要的是安排好我的工作和生活。

题目：他觉得爱情很重要。

答案：√

Material 1: Creo que el amor es una parte hermosa de la vida, pero para mí lo más importante es organizar mi trabajo y mi vida.

Pregunta: Él piensa que el amor es muy importante.

Respuesta: √

听力材料 2：在旅行中，我们要时刻注意安全，尤其是在陌生的地方，不要随便相信陌生人。

题目：在旅行中不需要注意安全。

答案：×

Material 2: Al viajar, siempre debemos prestar atención a la seguridad, especialmente en lugares desconocidos, y no confiar en los extraños a la ligera.

Pregunta: No es necesario prestar atención a la seguridad mientras viaja.

Respuesta: ×

听力材料 3: 我总是按时完成工作, 因为我知道时间很宝贵, 不能浪费。

题目: 他经常迟到。

答案: ×

Material 3: Siempre termino mi trabajo a tiempo porque sé que el tiempo es precioso y no se puede desperdiciar.

Pregunta: Suele llegar tarde.

Respuesta: ×

听力材料 4: 按照老师的安排, 我们这个学期要完成一份大作业, 需要分组合作。

题目: 他们没有大作业要完成。

答案: ×

Material 4: Según lo dispuesto por el maestro, tenemos que completar una tarea importante este semestre y debemos trabajar en grupos.

Pregunta: No tienen grandes tareas que hacer.

Respuesta: ×

听力材料 5: 我按照自己的计划安排了一次浪漫的约会, 但是天公不作美, 下着大雨。

题目: 他的约会很顺利。

答案: ×

Material 5: Organicé una cita romántica según mi plan, pero el clima no acompañaba y estaba lloviendo mucho.

Pregunta: Su cita salió bien.

Respuesta: ×

En primer lugar, analizamos el contenido de los textos. Los explicaremos uno por uno de forma escrita según los criterios de la rúbrica. Finalmente, realizaremos un análisis general y un resumen de todos los materiales generados por ChatGPT 3.5 en un gráfico de rúbrica.

**-Tipo de texto:** [4. Excelente] Puede comprender bien las instrucciones que ingresamos y generar materiales que cumplan con los requisitos y presentarlos con claridad.

- **Longitud del texto:** [3. Bueno] La mayoría de las oraciones del texto tienen la longitud requerida, es decir, de 30 a 60 caracteres, pero el material 3 “*Wo zongshi anshi wancheng gongzuo, yinwei wo zhidao shijian hen baogui, bu neng langfei.* (我总是按时完成工作，因为我知道时间很宝贵，不能浪费。)” [Siempre termino mi trabajo a tiempo porque sé que el tiempo es precioso y no se puede desperdiciar.]” sólo tiene 26 caracteres.

- **Adecuación al nivel indicado:** [1. Insuficiente] Aunque cada material incluye alguna de las palabras indicadas en las instrucciones, creemos que tanto el resto del vocabulario como las estructuras gramaticales generales son inferiores al nuevo HSK Nivel 4. Por ejemplo, el adverbio de grado “*zui (最)* [más]” en el material 1, la conjunción “*yinwei (因为)* [porque]” y el adverbio de tiempo “*zongshi (总是)* [siempre]” en el material 3 son puntos gramaticales del nivel 2 del HSK. Las estructuras de las oraciones son relativamente simples y no hay estructuras más avanzadas que permitan evaluar el nivel HSK 4. Sin embargo, hay que destacar que el material 5 utiliza el modismo “*tiangong bu zuo mei (天公不作美)* [Dios no es bueno en eso]”, que es un poco más difícil que el nivel 4 de nuevo HSK. Y la palabra “*mosheng (陌生)* [desconocido]” en el material 2 es del HSK nivel 5. Por lo tanto, en general, no hay puntos gramaticales o léxicos que reflejen lo que se espera en el HSK4.

- **Corrección a nivel gramatical y ortográfico:** [3. Bueno] La mayoría de las oraciones son gramaticalmente correctas, sin errores ortográficos y se expresan de manera clara. Identificamos una excepción en el material 5: dice “*xiazhe dayu (下着大雨)* [Está lloviendo fuertemente]”, aunque lo correcto en este contexto sería “*xiale dayu (下了大雨)* [Llovió mucho]”.

- **Naturalidad de la lengua y coherencia lógica del mensaje:** [2. Aceptable] El contenido del material es relativamente fluido y natural, pero hay algunas frases que creemos que tienen problemas lógicos. Por ejemplo, el Material 1 dice que “el amor es parte de la vida, pero lo más importante es organizar el trabajo y la vida”. Lo

primero y lo segundo parecen no tener lógica ni coherencia. Si se cambia la frase posterior a “*Danshi duiwo laishuo, zui zhongyao de shi gongzuo.* (但是对我来说, 最重要的是工作。 ) [Pero para mí lo más importante es el trabajo.]”, podría ser más coherente.

- **Autenticidad del texto:** [4. Excelente] Los temas de los materiales son relativamente reales e incluyen la vida, los viajes, el trabajo, el estudio y otros temas relativamente cercanos a la vida diaria.

En segundo lugar, analizamos los ejercicios diseñados por ChatGPT 3.5.

- **Tipo de ejercicio:** [4. Excelente] El contenido del material imita correctamente las preguntas de ejemplo, genera cinco preguntas de verdadero y falso, proporciona respuestas correctas y tiene un formato claro.

- **Relación al texto:** [3. Bueno] La mayoría de las preguntas están diseñadas para estar relacionadas con el contenido del material de comprensión oral, pero hay una pregunta que no tiene nada que ver con el contenido del dicho material. El material 3 trata sobre “*anshi wancheng gongzuo* (按时完成工作) [Completar el trabajo a tiempo], pero la pregunta dice que “*ta jingchang chidao* (他经常迟到) [Siempre llega tarde]”, no tiene nada que ver con el contenido del texto y el diseño no es razonable.

- **Adecuación al nivel indicado:** [4. Excelente] El contenido de las preguntas es fácil de entender, todas son oraciones cortas y la dificultad del vocabulario y la gramática es menor que la del contenido del texto. Por ejemplo, la palabra “*xuyao* (需要) [Necesitar]” en la pregunta del material 2, “*jingchang* (经常) [A menudo]” y “*chidao* (迟到) [Llegar tarde]” en la pregunta del material 3 son palabras del HSK 3, y la palabra “*shunli* (顺利) [Suavemente]” en la pregunta del material 5 es vocabulario del HSK 4. El adverbio negativo “*Bu* (不) [No]” en el material 2 y la frase con “*Meiyou* (没有) [No hay]” como en el material 4 son puntos gramaticales relativamente básicos del HSK 1.

- **Diseño de las preguntas:** [2. Aceptable] Algunas preguntas no están bien diseñadas. Aunque en el material 1 se dice que el amor es una parte hermosa de la vida, no menciona directamente la importancia del amor. Por lo tanto, creemos que establecer la pregunta como “*Ta jue de aiqing zui zhongyao* (他觉得爱情最重要) [Él piensa que el amor es lo más importante]” sería más fácil para los estudiantes tomar una decisión basada en el contenido del material. Por otra parte, la pregunta del material 2 se puede responder con sentido común.

- **Diseño de las opciones y respuestas:** [2. Aceptable] Solo una de las cinco preguntas tiene una respuesta de “correcta” y la distribución de respuestas correctas e incorrectas es desigual.

### 5.1.2 Análisis de los textos y ejercicios de elección múltiple generado por ChatGPT 3.5 (Ejemplo 2)

Segundo, mostramos los resultados de generación de las preguntas de elección múltiple en la segunda parte de comprensión oral del examen del nuevo HSK Nivel 4.

#### **Extracto 2. Ejercicios de elección múltiple por ChatGPT 3.5**

##### **PROMPT:**

听力材料:

女: 该加油了, 去机场的路上有加油站吗? 男: 有, 你放心吧。问: 男的主要是什么意思? 题目: A 去机场 B 快到了 C 油是满的 D 有加油站  
答案: D 有加油站。这是一道新 HSK 四级考试的听力材料和题目, 请你模仿它, 生成 5 道新 HSK4 级听力题。材料中分别包含以下词汇: 爱情、安排、安全、按时、按照。

Material para escuchar: Mujer: Es hora de repostar. ¿Hay alguna gasolinera de camino al aeropuerto? Hombre: Sí, no te preocupes. Pregunta: ¿Qué significa principalmente lo que dice el hombre? Opciones: A) Que van al aeropuerto B) Que están a punto de llegar C) Que el depósito está lleno D) Que hay una gasolinera Respuesta: D) Hay una gasolinera. Este es el material auditivo y las preguntas para el nuevo examen HSK Nivel 4. Imítalo y genera 5 nuevas preguntas de comprensión oral para HSK Nivel 4. El material contiene las siguientes palabras: *aiqing* (爱情) [amor], *anpai* (安排) [arreglar], *anquan* (安全) [seguridad], *anshi* (按时) [puntualmente] y *anzhao* (按照) [de acuerdo con].

##### **ChatGPT:**

听力材料 1:

女: 我们的婚礼安排得怎么样了?

男: 别担心, 我已经按照计划安排好了。

问: 男的主要是什么意思?

A. 计划有变化 B. 已经按计划安排好了 C. 还在考虑安排 D. 安排混乱

答案: B

Material 1:

Mujer: ¿Cómo van los preparativos de nuestra boda?

Hombre: No te preocupes, lo he arreglado según el plan.

Pregunta: ¿Qué significa principalmente el hombre?

A) Hay un cambio en el plan B) Ya está arreglado según lo planeado C) Aún está considerando el arreglo D) El arreglo es caótico

Respuesta: B

听力材料 2:

男: 你今晚会按时回家吗?

女: 当然, 我会注意安全的。

问: 女的主要是什么意思?

A. 不会按时回家 B. 会按时回家 C. 不在意安全 D. 担心安全问题

答案: B

Material 2:

Hombre: ¿Llegarás a casa a tiempo esta noche?

Mujer: Por supuesto, prestaré atención a la seguridad.

Pregunta: ¿Qué significa principalmente la mujer?

A) No regresará a casa a tiempo B) Volverá a casa a tiempo C) No le importa la seguridad D) Se preocupa por cuestiones de seguridad

Respuesta: B

听力材料 3:

女: 我们的爱情是不是太按照套路发展了?

男: 我觉得我们可以更自然一些, 不必那么拘泥于安排。

问: 男的主要是什么意思?

A. 赞同按照套路发展 B. 反对按照套路发展 C. 不确定是否按照套路发展 D. 无法评价按照套路发展

答案: B

Material 3:

Mujer: ¿Nuestro amor se basa demasiado en la rutina?

Hombre: Creo que podemos ser más naturales y no ser tan rígidos.

Pregunta: ¿Qué significa principalmente el hombre?

A) Está de acuerdo seguir las rutinas B) Se opone a seguir las rutinas C) No está seguro de si seguir las rutinas D) No puede evaluar el desarrollo del amor según las rutinas

Respuesta: B

听力材料 4:

男: 这个地方安全吗?

女: 是的, 他们按照国际标准建造的。

问: 女的主要是什么意思?

A. 不确定安全 B. 不安全 C. 安全 D. 不在意国际标准

答案: C

Material 4:

Hombre: ¿Es este lugar seguro?

Mujer: Sí, lo construyeron según estándares internacionales.

Pregunta: ¿Qué significa principalmente la mujer?

A) No está segura de la seguridad B) No, es inseguro. C) Sí, es seguro. D) No le importan los estándares internacionales

Respuesta: C

听力材料 5:

女: 今天约会的地点定好了, 准时到哦。

男: 放心, 我会按照你的要求准时到达的。

问: 男的主要是什么意思?

A. 不会按时到达 B. 会按时到达 C. 不在意要求 D. 不会准时到达

答案: B

Material 5:

Mujer: Ya hemos acordado el lugar de la cita de hoy. Por favor llega a tiempo.

Hombre: No te preocupes, llegaré a tiempo como me pides.

Pregunta: ¿Qué significa principalmente el hombre?

A) No llegará a tiempo B) Llegará a tiempo C) No le importa la petición de la mujer D) No llegará a tiempo

Respuesta: B

En primer lugar, analizamos el contenido de los textos.

- **Tipo de texto:** [4. Excelente] Comprende completamente las instrucciones y genera con precisión el formato requerido: una conversación entre un hombre y una mujer.

- **Longitud del texto:** [4. Excelente] Aunque no hay límite de caracteres aquí, ChatGPT imita muy bien el ejemplo proporcionado. Cada material es sólo una conversación, y el recuento de caracteres es aproximadamente de 20 a 50 palabras, lo que concuerda con el recuento de caracteres resumido en el artículo de Liang (2014: 8).

- **Adecuación al nivel indicado:** [1. Insuficiente] Algunas palabras son más difíciles que el HSK 4. La palabra “*hunli* (婚礼) [boda]” es de HSK 5 y “*taolu* (套路) [una serie de habilidades y trucos]” es una palabra que forma parte de la jerga de Internet, pero que no aparece en las listas de vocabulario de HSK. Por otra parte, “*juni* (拘泥) [limitarse]” y “*jianzao* (建造) [construir]” son vocabularios más allá del nivel 4 de HSK. Sin embargo, la dificultad gramatical es relativamente sencilla y no cumple con nuestras expectativas para el nivel 4 del HSK.

- **Corrección a nivel gramatical y ortográfico:** [4. Excelente] Todas las oraciones son gramaticalmente correctas, sin errores ortográficos y se expresan de manera clara.

- **Naturalidad de la lengua y coherencia lógica del mensaje:** [1. Insuficiente] La mayoría de los materiales tienen problemas de lógica o expresión rígida. En el material 2, el hombre le pregunta a la mujer si volverá a casa a tiempo, pero la mujer responde “*Wo hui zhuyi anquan de.* (我会注意安全的。)[Prestaré atención a la seguridad.]”, La respuesta es ilógica. El material 4 pregunta si este lugar es seguro, pero la respuesta es “*anzhao guoji biao zhun jianzao de* (按照国际标准建造的)[Construido según estándares internacionales]” Es más apropiado cambiar la pregunta a “*zhe dong dalou anquan ma?* (这栋大楼安全吗?) [¿Es seguro este edificio?]”. El material 5 incrusta la palabra “*anzhao* (按照) [de acuerdo con]” en la oración de respuesta. Según la palabra “*yuehui* (约会) [cita]” dicho por la mujer, se puede adivinar que los dos están en una relación, por lo que “*anzhao nide yaoqiu* (按照你的要求) [de acuerdo con tus requisitos]” es un poco forzado e inapropiado.

- **Autenticidad del texto:** [4. Excelente] Los textos son relativamente auténticos, la boda, quedar para una cita y el amor son temas cotidianos en la vida.

En segundo lugar, analizamos los ejercicios diseñados por ChatGPT 3.5.

- **Tipo de ejercicio:** [4. Excelente] ChatGPT imita correctamente las preguntas de ejemplo, genera cinco preguntas de elección múltiple, proporciona 4 opciones de cada pregunta y respuestas correctas.

- **Relación al texto:** [4. Excelente] El contenido de las preguntas está estrechamente integrado con el texto. Los estudiantes necesitan comprender el contenido del material primero para contestar las preguntas.

- **Adecuación al nivel indicado:** [4. Excelente] Las preguntas son fáciles de entender y no contienen palabras nuevas o estructuras gramaticales más complejas que el nivel de lenguaje del texto.

- **Diseño de las preguntas:** [2. Aceptable] Creemos que el diseño de las preguntas se puede mejorar. Quizás la pregunta de ejemplo era “*¿Nande zhuyao shi shenme yisi? (男的主要是什么意思?)* [¿Qué significa principalmente lo que dice el hombre?]", por lo que todas las preguntas generadas por ChatGPT preguntan qué significa lo que dice el personaje masculino o femenino. El diseño de la pregunta carece de innovación y diversidad. Para mejorar, la pregunta del material 3 se puede cambiar a “¿Cuál es la relación entre el hombre y la mujer?”

- **Diseño de las opciones y respuestas:** [3. Bueno] Las opciones se encuentran en el texto, no tienen significados repetidos entre ellas, hay sólo una respuesta correcta y no hay problemas gramaticales con las opciones. Pero en comparación con las opciones del documento de muestra del nuevo HSK 4, que presta más atención a la comprensión de diferentes vocabularios, las opciones generadas por ChatGPT son generalmente variaciones de la misma opción, como de afirmativa a negativa.

### **5.1.3 Análisis de los textos y ejercicios de elección múltiple generado por ChatGPT 3.5 (Ejemplo 3)**

Tercero, mostramos los resultados de generación de las preguntas de elección múltiple en la tercera parte de comprensión oral del examen del nuevo HSK Nivel 4.

### Extracto 3. Ejercicios de elección múltiple por ChatGPT 3.5

#### PROMPT:

听力文本:

男: 把这个材料复印五份, 一会儿拿到会议室发给大家。女: 好的。会议是下午三点吗? 男: 改了。三点半, 推迟了半个小时。女: 好, 602 会议室没变吧? 男: 对, 没变。题目: 会议几点开始? 选项: A 14: 00 B 15: 00 C 15: 30 D 18: 00 答案: C 这是一道新 HSK 四级考试的听力材料和题目, 请你模仿它, 生成 5 道新 HSK4 级听力题。听力材料为一男一女之间的 4-5 句对话, 字数一般为 30-60 字左右。材料 1 包含词汇: 爱情, 材料 2 包含词汇: 安排, 材料 3 包含词汇: 安全, 材料 4 包含词汇: 按时, 材料 5 包含词汇: 按照。

Texto para escuchar:

Hombre: Haz cinco copias de este material y llévalo a la sala de conferencias dentro de unos minutos para repartirlo entre todos. Mujer: De acuerdo. ¿La reunión es a las tres de la tarde? Hombre: Se ha cambiado. Se ha pospuesto media hora. Mujer: OK. La sala de conferencias 602 no ha cambiado, ¿verdad? Hombre: Sí, no ha cambiado. Pregunta: ¿A qué hora empieza la reunión? Opciones: A) A las dos B) A las tres C) A las tres y media D) A las seis Respuesta: C) A las tres y media Este es el material auditivo y las preguntas para el nuevo examen HSK Nivel 4. Imítalo y genera 5 nuevas preguntas de comprensión oral para HSK Nivel 4. Los materiales son de 4 a 5 oraciones de diálogo entre un hombre y una mujer, y tienen de 30 a 60 caracteres. El material 1 contiene la palabra: *aiqing* (爱情) [amor], el material 2 contiene la palabra: *anpai* (安排) [arreglar], el material 3 contiene la palabra: *anquan* (安全) [seguridad], el material 4 contiene la palabra: *anshi* (按时) [puntualmente] y el material 5 contiene la palabra: *anzhao* (按照) [de acuerdo con].

#### ChatGPT:

听力文本 1:

男: 亲爱的, 这周六我们可以一起去看电影吗?

女: 好啊, 几点开始?

男: 按照电影院的时间表, 下午三点半放映。

女: 好的, 我记住了, 我们在哪个影院?

男: 还是在我们常去的那家, 不变。

题目: 他们准备什么时候去看电影?

选项: A 下午三点 B 下午四点 C 下午五点 D 下午六点

答案: A

Material 1:

Hombre: Cariño, ¿podemos ir al cine juntos este sábado?

Mujer: Claro, ¿a qué hora empieza?

Hombre: Según el horario del cine, la proyectan a las tres y media de la tarde.

Mujer: Vale, ya me acuerdo, ¿en qué cine estamos?

Hombre: En el mismo de siempre, sin cambios.

Pregunta: ¿A qué hora van a ver la película?

Opción: A) A las tres de la tarde B A las cuatro de la tarde C A las cinco de la tarde D) A las seis de la tarde.

Respuesta: A

听力文本 2:

男: 小张, 这个月底有个重要的客户来访, 我们需要提前安排好。

女: 明白, 客户来的具体时间是?

男: 按照日程安排, 下午两点半到达。

女: 好的, 我会在那时候准备好接待的。

男: 非常感谢。

题目: 客户预计几点到达?

选项: A 下午一点 B 下午两点 C 下午两点半 D 下午三点

答案: C

Material 2:

Hombre: Xiao Zhang, hay una visita importante de un cliente a finales de este mes y tenemos que organizarla con antelación.

Mujer: Entendido, ¿cuál es la hora exacta de la visita del cliente?

Hombre: Llega a las dos y media de la tarde según el horario previsto.

Mujer: De acuerdo, estaré lista para recibirla a esa hora.

Hombre: Muchas gracias.

Pregunta: ¿A qué hora se espera la llegada del cliente?

Opción: A) A la una B) A las dos C) A las dos y media D) A las tres

Respuesta: C

听力文本 3:

男: 小心驾驶, 天气不太好。

女: 嗯, 我知道。按照路况, 应该会晚点到。

男: 安全第一, 慢慢开。

女: 好的, 我会注意的。

题目: 女士预计什么时候到达?

选项: A 早上九点 B 下午两点 C 下午三点 D 下午四点

答案: C

Material 3:

Hombre: Conduce con cuidado, el tiempo no es muy bueno.

Mujer: Bueno, ya lo sé. Según el estado de la carretera, llegaremos tarde.

Hombre: La seguridad es lo primero, conduce despacio.

Mujer: Sí, prestaré atención.

Pregunta: ¿A qué hora espera llegar la señora?

Opción: A) A las nueve de la mañana B) A las dos de la tarde C) A las tres de la tarde D) A las cuatro de la tarde

Respuesta: C

听力文本 4:

男: 明天的会议时间有变动吗?

女: 是的, 按时更新的日程, 会议推迟了半个小时。

男: 那应该是几点开始?

女: 原定的三点, 改为三点半。

男: 好的, 我会通知其他人。

题目: 会议改到什么时候开始?

选项: A 14: 00 B 15: 00 C 15: 30 D 16: 00

答案: C

Material 4:

Hombre: ¿Hay algún cambio en la hora de la reunión de mañana?

Mujer: Sí, la reunión se ha retrasado media hora según el horario actualizado puntualmente.

Hombre: ¿Y a qué hora debería empezar?

Mujer: La hora original de las tres se ha cambiado a las tres y media.

Hombre: De acuerdo, avisaré a los demás.

Pregunta: ¿A qué hora se ha reprogramado el comienzo de la reunión?

Opción: A) A las dos de la tarde B) A las tres de la tarde C) A las tres y media de la tarde D) A las cuatro de la tarde

Respuesta: C

听力文本 5:

男: 这个方案需要按照客户的需求来调整。

女: 明白, 那我们要多久完成?

男: 按照计划, 两个星期应该可以。

女: 好的, 我会尽快开始安排。

题目: 完成这个方案预计需要多长时间?

选项: A 一周 B 十天 C 两周 D 三周

答案: C

Material 5:

Hombre: Este programa debe adaptarse a las necesidades del cliente.

Mujer: Entendido, ¿cuánto tiempo tenemos para completarlo?

Hombre: Según el plan, dos semanas deberían estar bien.

Mujer: De acuerdo, empezaré a hacer los preparativos lo antes posible.

Pregunta: ¿Cuánto tiempo se prevé que tardaremos en completar este programa?

Opción: A) Una semana B) Diez días C) Dos semanas D) Tres semanas

Respuesta: C

En primer lugar, analizamos el contenido de los textos.

- **Tipo de texto:** [3. Bueno] Genera con precisión el formato requerido, pero el material 1 no contiene la palabra “*aiqing* (爱情) [amor]”.

- **Longitud del texto:** [4. Excelente] Cada pieza de material es un diálogo de 4-5 frases de longitud moderada, y aunque el *prompt* pide unas 30-60 caracteres, el ejemplo proporcionado tiene en realidad 73 caracteres, y el material final generado está en el rango de 55-85 caracteres, lo que pensamos que sigue estando en línea con la longitud del estilo que esperábamos conseguir.

- **Adecuación al nivel indicado:** [1. Insuficiente] Hay algo de gramática y vocabulario para el nivel HSK4, como “*jihua* (计划) [plan]” y “*ganxie* (感谢) [agradecer]”, y complementos de cantidad como “*Huiyi tuichi le bange xiaoshi.* (会议推迟了半个小时。)[La reunión se pospuso media hora.]”. Pero la mayor parte del vocabulario pertenece al nivel HSK 5, como “*fang'an* (方案) [plan]” “*richeng* (日程) [horario]” “*jiedai* (接待) [recibir]” “*jiashi* (驾驶) [conducir]” “*gengxin* (更新) [renovar]” o al nivel 6, como “*kehu* (客户) [cliente]” cliente, que no está adecuado al nivel HSK4.

- **Corrección a nivel gramatical y ortográfico:** [3. Bueno] La mayoría de las oraciones son gramaticalmente correctas y se expresan de manera clara. Sin embargo, en el Material 4, el uso de “*anshi* (按时) [puntualmente]” es incorrecto y “*anshi gengxin de richeng* (按时更新的日程) [horario actualizado a tiempo]” tiene un error gramatical y es difícil de entender. Debe cambiarse a “*anzhao gengxin de richeng* (按照更新的日程) [según el horario actualizado]”.

- **Naturalidad de la lengua y coherencia lógica del mensaje:** [3. Bueno] La mayoría del lenguaje es natural y lógica. Por ejemplo, palabras como “*haode* (好的) [vale]” “*wo zhidao* (我知道) [lo sé]” “*mingbai* (明白) [entender]” “*hao'a* (好啊) [sí]” “*en* (嗯) [de acuerdo]” se ajustan a las convenciones orales. Pero creemos que el material 2 omite información clave. La mujer preguntó al hombre el tiempo exacto a la que vino el cliente, pero el hombre sólo contestó qué hora era, no la fecha exacta.

- **Autenticidad del texto:** [4. Excelente] El texto es auténtico, se ajusta a los temas comunes del examen HSK, como el amor, el trabajo y el tráfico.

En segundo lugar, analizamos los ejercicios diseñados por ChatGPT 3.5.

- **Tipo de ejercicio:** [4. Excelente] ChatGPT imita correctamente las preguntas de ejemplo, genera cinco preguntas de elección múltiple, proporciona 4 opciones de cada pregunta y su respuesta. La forma es clara y concisa.

- **Relación al texto:** [4. Excelente] Las preguntas están estrechamente relacionadas con el texto y contribuyen a evaluar la comprensión del contenido.

- **Adecuación al nivel indicado:** [3. Bueno] La palabra “*yuji* (预计) [presuponer]” se encuentra en las preguntas de los Materiales 2, 3 y 5. Aunque la palabra no se encuentra en el vocabulario del HSK, podemos suponer que es difícil para los alumnos del nivel HSK 4, teniendo en cuenta que la palabra “*yuqi* (预期) [prever]” tiene un significado similar a la palabra “*yuji* (预计) [presuponer]”, que es una palabra del nivel HSK 6. No obstante, dado que esta palabra no tiene un impacto significativo en la comprensión total de la frase, le hemos dado una puntuación de 3.

- **Diseño de las preguntas:** [4. Excelente] Las preguntas están bien diseñadas, no se puede contestarse por sentido común o por coincidencia de léxico. El ChatGPT hace un buen trabajo imitando la pregunta sobre el tiempo en el ejercicio de muestra, que también se examinan en las cinco preguntas aquí.

- **Diseño de las opciones y respuestas:** [2. Aceptable] No hay opción correcta para el Material 1. El texto original dice “*xiawu sandian ban fangying* (下午三点半放映) [se proyecta a las tres y media de la tarde]”, pero la respuesta a la opción A) es “a las tres de la tarde”. Por lo tanto, la respuesta es incorrecta. En el Material 3, las cuatro opciones son todas horas precisas, pero el texto original sólo indica que “*yinggai hui wandian dao* (应该会晚点到) [debería llegar un poco tarde]”, sin especificar la hora, por lo que las opciones no tienen sentido. Sin embargo, en el Material 5, la respuesta correcta de “*liangge xingqi* (两个星期) [dos semanas]” se convierte en “*liangzhou* (两周) [dos semanas]”, que es un diseño flexible.

### 5.1.4 Análisis de la dicción oral generada por ChatGPT 3.5

Tras analizar los textos y ejercicios de comprensión oral generados por chatGPT 3.5, analizaremos a continuación la generación de la dicción oral de ChatGPT 3.5. En cuanto al contenido de la lectura en voz alta, pedimos a ChatGPT 3.5 que leyera el tercer material de opción múltiple que había generado (Véase Figura 1). A continuación mostraremos los resultados del análisis:

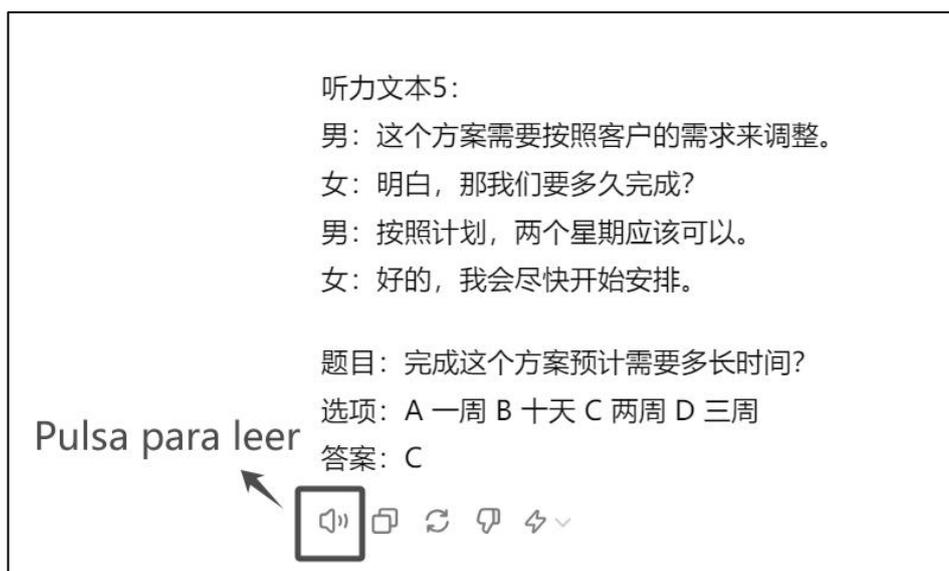


Figura 1. Función de lectura de texto de ChatGPT 3.5

- **Integridad del contenido:** [1. Insuficiente] Hay problemas de omisión y lectura múltiple, por ejemplo, la respuesta no se lee en el material 2, la opción A) no se lee en el material 4, y la primera frase dice dos veces la palabra “*huiyi* (会议) [reunión]”. Además los caracteres de “*nan* (男) [hombre]” y “*nü* (女) [mujer]” en el principio de cada material a veces se leen y a veces no. De todos modos, lo mejor sería que no leyera “*nan* (男) [hombre]” y “*nü* (女) [mujer]”, sino que utilizara voces de diferentes géneros para mostrar el diálogo tal com sería en condiciones de examen.

- **Precisión de la pronunciación:** [3. Bueno] La pronunciación es precisa y se ajusta al del mandarín estándar, incluyendo una correcta articulación de las consonantes iniciales y finales, las vocales y los tonos. Pero, la hora que se lee en el material 4 es incorrecta, las 14:00 se leen como “*shisi* (十四) [catorce]”, las 15:30 se leen como

“*shiwu sanshi* (十五三十) [quince treinta]”. En estos casos, lo más correcto sería añadir “*dian* (点) [hora]”, y leer las horas como “*shisi dian* (十四点) [las catorce horas]” y “*shiwu dian sanshi* (十五点三十) [las quince horas treinta]”.

- **Velocidad de habla:** [1. Insuficiente] La velocidad del habla es demasiado rápida y no se puede ajustar. Según la grabación de comprensión oral del HSK nivel 4 proporcionada por Zhongwen Kaoshi Fuwu Wang, son aproximadamente 3 caracteres por segundo, pero ChatGPT 3.5 lee aproximadamente 5 caracteres por segundo.

- **Cambio de tonos especiales:** [4. Excelente] En el material, la pronunciación de los diferentes tonos de “*yi* (一) [uno]” en “*yiqi* (一起) [juntos]” e “*yidian* (一点) [A la una]” es correcta, y la pronunciación de “*bubian* (不变) [no se cambia]” es justa. Como no hay casos de “*yi* (一) [uno]” con el cuarto tono en los materiales y no hay palabras con tonos neutros, verificamos la pronunciación introduciendo nuevos comandos en ChatGPT 3.5 y pidiéndole que leyera en voz alta las respuestas (Véase Extracto 4). Comprobamos que la pronunciación es acertada.

**Extracto 4. Dos palabras con un cuarto tono después del carácter “*yi* (一) [uno]” y dos palabras con tono neutro generadas por ChatGPT 3.5**

**PROMPT:**

请列举出 2 个 “一” 后面跟的是第四声字的词语。

Enumera 2 palabras en las que el carácter “*yi* (一) [uno]” vaya seguida de otra en cuarto tono.

**ChatGPT 3.5:**

以下是两个“一”后面跟第四声字的词语:

1. 一路
2. 一辆

Aquí hay dos palabras con un cuarto tono después del carácter “*yi* (一) [uno]”:

1. *yilu* (一路) [a lo largo del camino]
2. *yiliang* (一辆) [clasificador para modificar el coche]

**PROMPT:**

请你列举 2 个第二个字读轻声的词语。

Enumera dos palabras en las que la segunda se pronuncie con tono neutro.

**ChatGPT 3.5:**

以下是两个第二个字读轻声的词语：

1. 哥哥
2. 妹妹

Aquí hay dos palabras en las que la segunda se pronuncia con tono neutro:

1. *gege* (哥哥) [hermano]
2. *meimei* (妹妹) [hermana]

- **Entonación:** [4. Excelente] Puede transmitir con precisión las emociones y el tono del texto original, por ejemplo preguntas como “*Jidian kaishi?* (几点开始?) [¿Cuándo empieza?]” y énfasis como “*feichang ganxie* (非常感谢) [muchas gracias]”. Especialmente cuando se lee “*haode* (好的) [bueno]” y “*wo zhidao* (我知道) [lo sé]”, el tono es natural y vívido.

- **Naturalidad de habla:** [4. Excelente] La fluidez de habla es natural, con una conexión suave entre las oraciones, sin pausas o interrupciones evidentes, lo que hace que los oyentes sientan que están interactuando con una persona real.

- **Calidad de la voz:** [4. Excelente] El sonido es claro y melodioso, sin ruidos ni interrupciones, y ofrece opciones de 2 voces masculinas que se llaman “Ember” y “Cove” y femeninas que se llaman “Breeze” y “Juniper” para satisfacer las necesidades de diferentes usuarios.

- **Facilidad de la operación:** [1. Insuficiente] Primero, sólo se proporciona un botón de reproducción, sin función de pausa, por lo que cada vez que se pulsa, el material se vuelve a leer desde el principio. Segundo, como ChatGPT 3.5 genera 5 materiales de comprensión oral a la vez, los materiales se leen sucesivamente, por lo que no es posible diseñar tiempos de pausa para que los alumnos lean o escriban preguntas. Tercero, como se trata de un diálogo entre un hombre y una mujer, pero sólo se puede seleccionar uno de los géneros de voz para leer todo el texto, no es adecuado para la comprensión oral del diálogo. Por último, no es posible establecer la velocidad de los materiales de comprensión oral ni exportarlos. Por lo tanto, no es conveniente para los profesores.

## 5.2 Análisis de Gemini

A continuación, se presentará los materiales de comprensión oral y las preguntas generadas por Gemini utilizando las mismas instrucciones ingresadas en ChatGPT. Para alcanzar el texto final que consideramos evaluable, probamos varias instrucciones adicionales para optimización después de ingresar las instrucciones iniciales. Omitiremos el proceso aquí. Este proceso de entrenamiento de IA se detallará en el Capítulo 6.

### 5.2.1 Análisis de los textos y ejercicios de verdadero y falso generado por Gemini (Ejemplo 1)

En primer lugar, mostramos los resultados de generación de las preguntas de verdadero y falso en la primera parte de comprensión oral del examen del nuevo HSK Nivel 4. Dado que en algunos casos Gemini añade una explicación al final, también se ha incluido y se ha traducido del chino al español en este trabajo.

#### Extracto 5. Ejercicios de verdadero y falso generado por Gemini

##### PROMPT:

听力材料：现在我很少看电视，其中一个原因是，广告太多了，不管什么时间，也不管什么节目，只要你打开电视，总能看到那么多的广告，浪费我的时间。题目：他喜欢看电视广告。答案：×。这是一道新 HSK 四级考试的听力材料和题目，请你模仿它，生成 5 道听力题，听力材料 30 到 60 字，材料中分别包含以下词汇：爱情、安排、安全、按时、按照。题目答案有对有错。

Material para escuchar: ahora rara vez veo televisión. Una de las razones es que hay demasiados anuncios, no importa la hora o el programa, siempre que enciendas el televisor, siempre verás muchos anuncios, lo cual es un desperdicio de mi tiempo. Pregunta: Le gusta ver anuncios de televisión. Respuesta: ×. Este es un material de escucha y una pregunta para el nuevo examen HSK Nivel 4. Imítalo y genera 5 preguntas de comprensión oral. El material tiene de 30 a 60 caracteres. El material contiene las siguientes palabras: *aiqing* (爱情) [amor], *anpai* (安排) [arreglar], *anquan* (安全) [seguridad], *anshi* (按时) [puntualmente] y *anzhao* (按照) [de acuerdo con]. Hay respuestas correctas e incorrectas a las preguntas.

##### GEMINI:

题目 1:

听力材料:

小丽很喜欢看电影,她经常周末去电影院看电影。最近,她听说了一部新电影,非常想去看,于是她就提前在网上订好了电影票。

问题:小丽不喜欢看电影。(是/否)

答案:否

解释:文本中明确指出小丽很喜欢看电影,经常周末去电影院看电影,并提前在网上订好了新电影的票,因此答案为“否”。

Ejercicio 1:

Material de comprensión oral:

A Xiaoli le gusta mucho ver películas y suele ir al cine los fines de semana. Hace poco le hablaron de una nueva película y tenía muchas ganas de ir a verla, así que reservó las entradas por Internet con antelación.

Pregunta: A Xiaoli no le gusta ir al cine. (Sí/No)

Respuesta: No

Explicación: El texto deja claro que a Xiaoli le gusta ir al cine, suele ir al cine los fines de semana y ha reservado las entradas para una nueva película por Internet con antelación, por lo que la respuesta es “no”.

题目 2:

听力材料:

王先生是一位非常注重安全的人,他家里的所有电器都安装了漏电保护装置,他还经常检查家里的消防设施,确保家里安全。

问题:王先生不注重安全。(是/否)

答案:否

解释:文本中描述王先生采取了多项措施来确保家庭安全,包括安装漏电保护装置和定期检查消防设施,因此答案为“否”。

Ejercicio 2:

Material de comprensión oral:

El Sr. Wang es una persona muy preocupada por la seguridad. Todos los aparatos eléctricos de su casa están equipados con dispositivos de protección contra fugas a tierra, y comprueba regularmente las instalaciones de seguridad contra incendios de su casa para asegurarse de que es segura.

Pregunta: Al Sr. Wang no le preocupa la seguridad. (Sí/No)

Respuesta: No

Explicación: El texto describe que el Sr. Wang ha tomado una serie de medidas para garantizar la seguridad de su vivienda, como la instalación de dispositivos de protección contra fugas a tierra y la comprobación periódica de las instalaciones de seguridad contra incendios, por lo que la respuesta es “no”.

题目 3:

听力材料:

李丽是一位很守时的人,她总是提前几分钟到公司上班,从不迟到。她认为,守时是一种美德。

问题:李丽不守时。(是/否)

答案:否

解释：文本中强调李丽是一位守时的人，总是提前几分钟到公司上班，从不迟到，并认为守时是一种美德，因此答案为“否”。

Ejercicio 3:

Material de comprensión oral:

Li Li es una persona muy puntual, siempre llega al trabajo unos minutos antes y nunca llega tarde. Cree que la puntualidad es una virtud.

Pregunta: Li Li no es puntual. (Sí/No)

Respuesta: No

Explicación: El texto destaca que Li Li es una persona puntual que siempre llega al trabajo unos minutos antes, nunca llega tarde y cree que la puntualidad es una virtud, por lo que la respuesta es “no”.

题目 4:

听力材料:

小明和小红是一对情侣，他们计划下个月去旅行。小明已经安排好了旅行计划，订好了机票和酒店，但小红却不太满意这个计划，她希望能够去一些她一直想去的地方。

问题：小红对旅行计划满意。（是/否）

答案：否

解释：文本中指出小红不太满意小明安排的旅行计划，希望能够去一些她一直想去的地方，因此答案为“否”。

Ejercicio 4:

Material de comprensión oral:

Xiao Ming y Xiao Hong son pareja y están planeando un viaje para el mes que viene. Xiao Ming ya ha organizado los planes de viaje y ha reservado los vuelos y los hoteles, pero Xiao Hong no está muy contenta con el plan, le gustaría ir a algunos lugares que siempre ha querido visitar.

Pregunta: Xiao Hong está contenta con los planes de viaje. (Sí/No)

Respuesta: No

Explicación: En el texto se dice que Xiaohong no está del todo satisfecha con los planes de viaje organizados por Xiaoming y que le gustaría ir a algunos lugares que siempre ha querido visitar, por lo que la respuesta es “no”.

题目 5:

听力材料:

张老师是一位经验丰富的老师，她总是按照教学计划上课，从不随意更改课程内容。她也经常给学生布置作业，并认真批改作业。

问题：张老师不按照教学计划上课。（是/否）

答案：否

解释:

文本中描述张老师是一位经验丰富的老师，总是按照教学计划上课，从不随意更改课程内容，因此答案为“否”。

Ejercicio 5:

Material de comprensión oral:

La profesora Zhang es una profesora experimentada que siempre sigue el plan de la lección y nunca cambia el contenido de la misma a su antojo. También suele asignar deberes a sus alumnos y los corrige cuidadosamente.

Pregunta: La profesora Zhang no sigue el plan de la lección. (Sí/No)

Respuesta: No

Explicación: El texto describe a la señora Zhang como una profesora experimentada que siempre sigue el plan de la lección y nunca cambia el contenido de la misma a su antojo, por lo que la respuesta es “no”.

En primer lugar, analizamos el contenido de los textos.

**-Tipo de texto:** [3. Bueno] Genera oraciones simples como materiales de comprensión oral según la instrucción, pero en el contenido solo aparecen 3 palabras: “*anpai* (安排) [arreglar]”, “*anquan* (安全) [seguridad]” y “*anzhao* (按照) [de acuerdo con]”, mientras que la palabra “*aiqing* (爱情) [amor]” y “*anshi* (按时) [puntualmente]” no se incluyen. Además, el material de ejemplo está narrado desde la perspectiva de primera persona, pero todos los materiales generados están narrados desde la perspectiva de tercera persona.

- **Longitud del texto:** [3. Bueno] La mayoría de las oraciones del texto tienen la longitud requerida, que tiene de 30 a 60 caracteres, pero el material 4 tiene 74 caracteres.

- **Adecuación al nivel indicado:** [2. Aceptable] La gramática y la dificultad de las oraciones se adecuan al nivel nuevo HSK 4. Pero hay algunas palabras que son más difíciles que el HSK 4, como “*xiaofang* (消防) [extinción de incendios]” “*quebao* (确保) [asegurarse]” y “*buzhi* (布置) [disponer]”, que son palabras del nivel HSK 6.

- **Corrección a nivel gramatical y ortográfico:** [4. Excelente] Todas las oraciones son gramaticalmente correctas, sin errores ortográficos y se expresan de manera clara.

- **Naturalidad de la lengua y coherencia lógica del mensaje:** [4. Excelente] Cada material utiliza conjunciones, como “*yushi* (于是) [Por lo tanto]” “*hai* (还) [Aún]”

“*dan* ( 但 ) [Pero]” y “*ye* ( 也 ) [Y]”. Las oraciones anteriores y posteriores están conectadas de forma natural y lógica.

- **Autenticidad del texto:** [4. Excelente] Los temas son relativamente cercanos a la vida diaria, incluyen ver películas, vida familiar, trabajo, viaje y estudio.

En segundo lugar, analizamos los ejercicios diseñados por Gemini.

- **Tipo de ejercicio:** [4. Excelente] El material imita adecuadamente las preguntas de ejemplo, genera cinco preguntas de verdadero o falso, proporciona las respuestas correctas y tiene un formato claro, además, ofrece las explicaciones.

- **Relación al texto:** [4. Excelente] Todas las preguntas están estrechamente relacionadas con el texto.

- **Adecuación al nivel indicado:** [4. Excelente] Las preguntas son fáciles de entender y no tienen puntos gramaticales o léxicos que superan el nivel del texto.

- **Diseño de las preguntas:** [1. Insuficiente] Consideramos que el diseño de las preguntas es demasiado simple, ya que las preguntas del material 1, 2 y 3 están todas relacionadas con la primera oración del texto. Los estudiantes pueden responder sin escuchar el resto del contenido. Además, en el ejemplo, “*Ta bu xihuan kan dianshi guanggao.* (她不喜欢看电视广告) [A él no le gustan los anuncios de televisión]” no se menciona explícitamente en el material, los estudiantes deben inferirlo a partir de “*henshao kan dianshi shi yinwei guanggao taiduo le* (很少看电视是因为广告太多了) [Ve muy poco la televisión porque hay demasiados anuncios]”. Sin embargo, todas las preguntas aquí están copiadas directamente del contenido original, sin evaluar la comprensión profunda y la capacidad de inferencia de los estudiantes.

- **Diseño de las opciones y respuestas:** [1. Insuficiente] Todas las preguntas tienen una respuesta de “falsa” y la distribución de respuestas correctas e incorrectas es desigual.

## 5.2.2 Análisis de los textos y ejercicios de elección múltiple generado por Gemini (Ejemplo 2)

Segundo, mostramos los resultados de generación de las preguntas de elección múltiple en la segunda parte de comprensión oral del examen del nuevo HSK Nivel 4.

### Extracto 6. Ejercicios de elección múltiple por Gemini

#### PROMPT:

听力材料:

女: 该加油了, 去机场的路上有加油站吗? 男: 有, 你放心吧。问: 男的主要是什么意思? 题目: A 去机场 B 快到了 C 油是满的 D 有加油站  
答案: D 有加油站。这是一道新 HSK 四级考试的听力材料和题目, 请你模仿它, 生成 5 道新 HSK4 级听力题。材料中分别包含以下词汇: 爱情、安排、安全、按时、按照。

Material para escuchar: Mujer: Es hora de repostar. ¿Hay alguna gasolinera de camino al aeropuerto? Hombre: Sí, no te preocupes. Pregunta: ¿Qué significa principalmente lo que dice el hombre? Opciones: A) Que van al aeropuerto B) Que están a punto de llegar C) Que el depósito está lleno D) Que hay una gasolinera Respuesta: D) Hay una gasolinera. Este es el material auditivo y las preguntas para el nuevo examen HSK Nivel 4. Imítelo y genere 5 nuevas preguntas auditivas HSK Nivel 4. El material contiene las siguientes palabras: *aiqing* (爱情) [amor], *anpai* (安排) [arreglar], *anquan* (安全) [seguridad], *anshi* (按时) [puntualmente] y *anzhao* (按照) [de acuerdo con].

#### GEMINI:

好的, 以下是五道新的 HSK4 级听力题, 材料中分别包含以下词汇: 爱情、安排、安全、按时、按照:

Bien, aquí hay cinco nuevas preguntas de comprensión oral de HSK nivel 4. Los materiales incluyen el siguiente vocabulario: *aiqing* (爱情) [amor], *anpai* (安排) [arreglar], *anquan* (安全) [seguridad], *anshi* (按时) [puntualmente] y *anzhao* (按照) [de acuerdo con]:

#### 题目 1:

听力材料:

女: 今天是我们结婚纪念日, 你想去哪里庆祝?

男: 我已经安排好了, 我们一起去浪漫的餐厅吃晚餐, 然后去看电影。

问题:

男的安排是什么?

答案:

A. 去浪漫的餐厅吃晚餐 B. 看电影 C. 庆祝结婚纪念日 D. 去机场

正确答案: A. 去浪漫的餐厅吃晚餐

**解释:**

男人说他已经安排好了他们去浪漫的餐厅吃晚餐, 然后去看电影, 这表明这部分是他计划的一部分。

Pregunta 1:

Materiales de comprensión oral:

Mujer: Hoy es nuestro aniversario de bodas. ¿Dónde quieres celebrarlo?

Hombre: Ya he reservado para que vayamos a cenar a un restaurante romántico y luego vayamos al cine.

Pregunta: ¿Cuál es el plan del hombre?

Respuesta: A) Ir a cenar a un restaurante romántico B) Ver una película C) Celebrar su aniversario de bodas D) Ir al aeropuerto

Respuesta correcta: A) Ir a cenar a un restaurante romántico.

Explicación: El hombre dice que había hecho reservas para que cenaran en un restaurante romántico y luego fueran al cine, sugiriendo que esto era parte de su plan.

**题目 2:**

**听力材料:**

男: 你总是按时完成任务, 真的很可靠。

女: 谢谢你的夸奖, 我会继续努力的。

**问题:**

女的是怎么样的?

**答案:**

A. 粗心大意 B. 不可靠 C. 总是按时完成任务 D. 不认真

**正确答案:** C. 总是按时完成任务

**解释:**

男人称赞女人总是按时完成任务, 这意味着她很可靠。

Pregunta 2:

Materiales de comprensión oral:

Hombre: Siempre completas las tareas a tiempo, eres realmente fiable.

Mujer: Gracias por tu cumplido. Seguiré trabajando duro.

Pregunta: ¿Cómo es la mujer?

Respuesta: A) Descuidada B) Poco fiable C) Siempre completa las tareas a tiempo D) No es seria

Respuesta correcta: C) Completar siempre las tareas a tiempo

Explicación: Cuando el hombre elogia a la mujer por completar siempre las tareas a tiempo, significa que es confiable.

**题目 3:**

**听力材料:**

女: 我很注重安全, 从来不闯红灯。

男: 你真是太棒了, 我们要一起遵守交通规则。

**问题:**

女的是怎么样的？

**答案：**

A. 总是闯红灯 B. 不注重安全 C. 很有爱心 D. 非常守法

**正确答案：** D. 非常守法

**解释：**

女人说她很注重安全，从来不闯红灯，这表明她很守法。

Pregunta 3:

Materiales de comprensión oral:

Mujer: Presto atención a la seguridad y nunca me paso un semáforo en rojo.

Hombre: Muy bien. Debemos obedecer las reglas de tráfico juntos.

Pregunta: ¿Cómo es la mujer?

Respuesta: A) Pasa siempre los semáforos en rojo B) No presta atención a la seguridad C) Muy atenta D) Muy respetuosa con la ley

Respuesta correcta: D) Muy respetuoso con la ley.

Explicación: La mujer dijo que presta atención a la seguridad y nunca se pasa un semáforo en rojo, lo que demuestra que es muy respetuosa con la ley.

**题目 4：**

**听力材料：**

男：你对我们的爱情还有信心吗？

女：当然有，我相信我们会一直幸福下去的。

**问题：**

女的是怎么样的？

**答案：**

A. 对爱情没有信心 B. 很伤心 C. 非常乐观 D. 很失望

**正确答案：** C. 非常乐观

**解释：**

女人说她相信他们的爱情，他们会一直幸福下去，这表明她对他们的关系很乐观。

Pregunta 4:

Materiales de comprensión oral:

Hombre: ¿Aún tienes confianza en nuestro amor?

Mujer: Por supuesto, creo que siempre seremos felices.

Pregunta: ¿Cómo es la mujer?

Respuesta: A) No tiene confianza en el amor B) Muy triste C) Muy optimista D) Muy decepcionada

Respuesta correcta: C) Muy optimista

Explicación: La mujer dice que cree en su amor y que siempre serán felices, lo que demuestra que es optimista sobre su relación.

**题目 5：**

**听力材料：**

老师：同学们，你们要按照教学计划复习，才能取得好成绩。

学生：好的，老师，我们会认真复习的。

**问题:**

老师希望学生做什么?

**答案:**

A. 提前预订电影票 B. 去机场 C. 按照教学计划复习 D. 庆祝结婚纪念日

**正确答案:** C. 按照教学计划复习

**解释:**

老师告诉学生要按照教学计划复习, 这表明这是他希望学生做的事情。

Pregunta 5:

Materiales de comprensión oral:

Maestro: Estudiantes, deben repasar según el plan de enseñanza para poder lograr buenos resultados.

Estudiante: Está bien, maestro, repasaremos detenidamente.

Pregunta: ¿Qué quiere el maestro que hagan los estudiantes?

Respuesta: A) Reservar entradas de cine con antelación B) Ir al aeropuerto C) Repasar según el plan de enseñanza D) Celebrar el aniversario de boda

Respuesta correcta: C) Revisar según el plan de enseñanza

Explicación: El profesor les dice a los estudiantes que revisen de acuerdo con el plan de enseñanza, lo que demuestra que eso es lo que quiere que hagan los estudiantes.

En primer lugar, analizamos el contenido de los textos.

- **Tipo de texto:** [4. Excelente] Comprende completamente las instrucciones y genera con precisión el formato requerido: un diálogo entre un hombre y una mujer. Además, se incluyen todas las 5 palabras proporcionadas en la instrucción.

- **Longitud del texto:** [4. Excelente] La extensión del texto de todos los materiales es de aproximadamente 30 a 50 palabras, lo cual es moderado y apropiado.

- **Adecuación al nivel indicado:** [3. Bueno] La mayoría de las palabras, la gramática y la dificultad de las oraciones se adecuan al nivel nuevo HSK 4. Excepto algunas palabras que son palabras de nivel HSK 5, como “*jinian* (纪念) [conmemorar]” “*qingzhu* (庆祝) [celebrar]” “*kekao* (可靠) [confiable]” y “*zunshou* (遵守) [obedecer]”.

- **Corrección a nivel gramatical y ortográfico:** [4. Excelente] Todas las oraciones son gramaticalmente correctas, sin errores ortográficos y se expresan de manera clara.

- **Naturalidad de la lengua y coherencia lógica del mensaje:** [4. Excelente] El lenguaje es fluido y local, el diálogo es natural y las frases anteriores y posteriores están conectadas.

- **Autenticidad del texto:** [4. Excelente] Los textos son relativamente auténticos y cuentan con temas cordiales, como aniversario de boda, trabajo, moral, amor y estudio.

En segundo lugar, analizamos los ejercicios diseñados por Gemini.

- **Tipo de ejercicio:** [3. Bueno] Gemini imita correctamente las preguntas de ejemplo, genera cinco preguntas de elección múltiple, proporciona 4 opciones de cada pregunta y respuestas correctas. Sin embargo, la forma en que se presenta el contenido no es muy concisa y el título de opciones está escrito como “*da'an* (答案) [respuesta]”. Es más exacto escribirlo como “*xuanxiang* (选项) [opciones]” aquí.

- **Relación al texto:** [4. Excelente] Las preguntas están profundamente relacionadas con el texto. Los estudiantes deben entender el contenido del material primero para poder responder a las preguntas.

- **Adecuación al nivel indicado:** [3. Bueno] Las preguntas son generalmente comprensibles, pero en algunas opciones hay modismos como “*cuxin-dayi* (粗心大意) [descuidado]”, palabras que superan el nivel del HSK 4 como “*shoufa* (守法) [atender a la ley]” “*leguan* (乐观) [optimismo]”, puede resultar difícil para los estudiantes.

- **Diseño de las preguntas:** [2. Aceptable] Las preguntas del material 1 y 5 tienen un significado más natural y claro, pero el lenguaje de los materiales 2, 3 y 4 es un poco rígido, lo que puede afectar a la comprensión de los estudiantes. La pregunta del material 2 se puede cambiar por “*Nüde shi shenme yang de ren?* (女的是什么样的人?) [¿Qué tipo de persona es la mujer?]”, y el material 4 se puede cambiar por “*Nüde shi shenme yang de xinqing?* (女的是什么样的心情?) [¿En qué estado de ánimo está la mujer?]”.

- **Diseño de las opciones y respuestas:** [2. Aceptable] La mayoría de las respuestas se pueden encontrar en el texto original y un pequeño número de preguntas requieren que los estudiantes hagan inferencias, como el material 3, material 4. También se pueden generar distractores relacionadas con la opción correcta, lo que puede examinar el dominio de los estudiantes de vocabularios múltiples. El diseño de la opción sigue siendo relativamente bueno. Sin embargo, la respuesta correcta en el material 1 no es sólo A), sino también B) que se menciona en el texto original, por lo que la respuesta aquí no es única. El material 5 se puede responder directamente a través de eliminar las opciones, porque “*qu jichang* (去机场) [ir al aeropuerto]” y “*qingzhu jiehun jinian ri* (庆祝结婚纪念日) [celebrar aniversarios de boda]” no encajan en los temas habituales de profesores y alumnos.

### 5.2.3 Análisis de los textos y ejercicios de elección múltiple generado por Gemini (Ejemplo 3)

Tercero, mostramos los resultados de generación de las preguntas de elección múltiple en la tercera parte de comprensión oral del examen del nuevo HSK Nivel 4. Cabe mencionar que inicialmente Gemini no proporcionaba la respuesta a cada pregunta, pero a efectos del análisis, hemos combinado los comandos de entrada posteriores para que la respuesta proporcionada por Gemini coincida con la pregunta. En el capítulo 6 mostramos el proceso.

#### Extracto 7. Ejercicios de elección múltiple por Gemini

##### PROMPT:

听力文本:

男: 把这个材料复印五份, 一会儿拿到会议室发给大家。女: 好的。会议是下午三点吗? 男: 改了。三点半, 推迟了半个小时。女: 好, 602 会议室没变吧? 男: 对, 没变。题目: 会议几点开始? 选项: A 14: 00 B 15: 00 C 15: 30 D 18: 00 答案: C 这是一道新 HSK 四级考试的听力材料和题目, 请你模仿它, 生成 5 道新 HSK4 级听力题。听力材料为一男一女之间的 4-5 句对话, 字数一般为 30-60 字左右。材料 1 包含词汇: 爱情, 材料 2 包含词汇: 安排, 材料 3 包含词汇: 安全, 材料 4 包含词汇: 按时, 材料 5 包含词汇: 按照。

Texto para escuchar:

Hombre: Haz cinco copias de este material y llévalo a la sala de conferencias dentro de unos minutos para repartirlo entre todos. Mujer: De acuerdo. ¿La reunión es a las tres de la tarde? Hombre: Se ha cambiado. Se ha pospuesto media hora. Mujer: OK. La sala de conferencias 602 no ha cambiado, ¿verdad? Hombre: Sí, no ha cambiado. Pregunta: ¿A qué hora empieza la reunión? Opciones: A) A las dos B) A las tres C) A las tres y media D) A las seis Respuesta: C) A las tres y media Este es el material auditivo y las preguntas para el nuevo examen HSK Nivel 4. Imítalo y genera 5 nuevas preguntas de comprensión oral para HSK Nivel 4. Los materiales son de 4 a 5 oraciones de diálogo entre un hombre y una mujer, y tienen de 30 a 60 caracteres. El material 1 contiene la palabra: *aiqing* (爱情) [amor], el material 2 contiene la palabra: *anpai* (安排) [arreglar], el material 3 contiene la palabra: *anquan* (安全) [seguridad], el material 4 contiene la palabra: *anshi* (按时) [puntualmente] y el material 5 contiene la palabra: *anzhao* (按照) [de acuerdo con].

### GEMINI:

材料一：爱情

对话

男：你好，请问你喜欢什么样的爱情？

女：我喜欢平淡简单的爱情，两个人互相理解，互相支持，一起面对生活中的风雨。

男：你的想法和我一样。我认为爱情不需要轰轰烈烈，只要两个人真心相爱，携手共度一生就足够了。

问题：女士喜欢什么样的爱情？

- A. 轰轰烈烈的爱情
- B. 平淡简单的爱情
- C. 充满挑战的爱情
- D. 浪漫唯美的爱情

（这里我们再次输入指令，要求 Gemini 提供问题的答案）

答案：B

女士喜欢平淡简单的爱情，认为两个人互相理解、互相支持，一起面对生活中的风雨才是最重要的。

Material 1: Amor

Diálogo

Hombre: Hola, ¿qué tipo de amor te gusta?

Mujer: Me gusta el amor simple y llano, dos personas que se entienden, se apoyan y afrontan juntas las tormentas de la vida.

Hombre: Tu idea es la misma que la mía. Creo que el amor no necesita ser sensacional, mientras dos personas se amen sinceramente es suficiente de la mano para pasar la vida de la mano.

Pregunta: ¿Qué tipo de amor le gusta a una señora?

- A) Amor sensacional B) Amor sencillo C) Amor desafiante D) Amor romántico

(Aquí introducimos de nuevo el comando, pidiendo a Gemini que proporcione la respuesta a la pregunta)

Respuesta: B

La señora prefiere el amor sencillo y piensa que lo más importante es que dos personas se entiendan y se apoyen mutuamente y afronten juntas las tormentas de la vida.

材料二：行程

对话

女：你好，请问下周你有时间吗？我想和你一起安排一下下个月的旅行计划。

男：下周我比较忙，要参加公司的一个培训项目。不过下个月我应该有时间，你想去哪里旅行？

女：我一直想去欧洲，听说那里的风景很美，文化也很丰富。

男：欧洲确实是一个不错的选择，我们可以去法国、意大利、西班牙等国家玩一圈。

问题：下个月女士和男士想去哪里旅行？

A. 欧洲

B. 亚洲

C. 美洲

D. 非洲

答案：A

下个月女士和男士想去欧洲旅行。他们觉得欧洲风景优美，文化丰富，可以去法国、意大利、西班牙等国家游玩。

Material 2: Itinerario

Diálogo

Mujer: Hola, ¿tienes tiempo la semana que viene? Me gustaría organizar con usted un plan de viaje para el mes que viene.

Hombre: La semana que viene estoy bastante ocupado, tengo que asistir a un programa de formación en mi empresa. ¿Dónde quieres viajar?

Mujer: Siempre he querido ir a Europa, he oído que los paisajes son preciosos y la cultura muy rica.

Hombre: Europa es realmente una buena opción, podemos ir a Francia, Italia, España y otros países para hacer un *tour*.

Pregunta: ¿Adónde quieren viajar la señora y el señor el mes que viene?

A) Europa B) Asia C) América D) África

Respuesta: A) Europa

La señora y el señor quieren viajar a Europa el mes que viene. Piensan que Europa tiene bellos paisajes y una rica cultura, por lo que pueden visitar países como Francia, Italia y España.

材料三：安全

对话

男：你好，请问你知道怎么使用灭火器吗？

女：我知道一些基本的步骤。首先，我们需要拉开灭火器的保险栓，然后瞄准火源，最后用力按压喷射口。

男：很好，你记得很清楚。在使用灭火器之前，我们还应该注意一些什么？

女：我们应该先远离火源，并拨打 119 报警电话。

问题：使用灭火器之前应该注意什么？

- A. 拉开保险栓，瞄准火源，用力按压喷射口
- B. 拨打 119 报警电话
- C. 远离火源
- D. 检查灭火器是否有效

答案：C

使用灭火器之前，应该先远离火源，并拨打 119 报警电话。这可以确保人身安全，并提高灭火效率。

### Material 3: Seguridad

#### Diálogo

Hombre: Hola, ¿sabes utilizar un extintor?

Mujer: Conozco algunos pasos básicos. En primer lugar, hay que tirar del pasador de seguridad del extintor, luego apuntar al foco del incendio y, por último, presionar con fuerza el chorro.

Hombre: Bien, lo recuerdas bien. ¿A qué más debemos prestar atención antes de utilizar el extintor?

Mujer: Primero debemos alejarnos del foco del incendio y llamar a la 119 policía.

Pregunta: ¿A qué debemos prestar atención antes de utilizar un extintor?

A) Quitar el seguro, apuntar al foco del incendio y presionar con fuerza la boquilla pulverizadora. B) Llamar a la policía al 119 C) Alejarse del foco del incendio D) Comprobar si el extintor es eficaz

Respuesta: C

Antes de utilizar un extintor, debe alejarse del foco del incendio y llamar al 119. Esto garantiza la seguridad personal y aumenta la eficacia de la extinción del fuego.

### 材料四：准时

#### 对话

男：明天早上八点开会，你记得按时参加吗？

女：当然记得，我已经把会议时间记在我的日程表上了。我会提前十分钟到达会议室。

男：很好，你做事很认真负责。会议的主要内容是讨论下个月的销售计划，你有什么想法吗？

女：我认为我们应该重点关注新产品的推广，同时也要加强对老客户的维护。

问题：女士将如何准备会议？

- A. 提前十分钟到达会议室
- B. 提前阅读会议材料
- C. 准备发言稿

#### D. 做好会议记录

答案：A

女士将提前十分钟到达会议室，以确保准时参加会议。她做事认真负责，并提前做好了准备。

#### Material 4: Puntualmente

##### Diálogo

Hombre: La reunión se celebrará mañana a las ocho de la mañana, ¿te acuerdas de ser puntual?

Mujer: Claro que me acuerdo, he anotado la hora de la reunión en mi agenda. Llegaré a la sala de reuniones diez minutos antes.

Hombre: Bien, eres muy seria y responsable. El contenido principal de la reunión es discutir el plan de ventas del próximo mes, ¿tienes alguna idea?

Mujer: Creo que deberíamos centrarnos en la promoción de nuevos productos y también reforzar el mantenimiento de nuestros antiguos clientes.

Pregunta: ¿Cómo se preparará la señora para la reunión?

A) Llegará a la sala de reuniones con diez minutos de antelación B) Leerá el material de la reunión con antelación C) Preparará un discurso D) Levantará el acta de la reunión

#### 材料五：按照

##### 对话

男：你好，请问你知道如何按照食谱做饭吗？

女：我知道一些基本的烹饪技巧。我可以按照食谱上的步骤来做饭，但有时候我也会根据自己的喜好进行一些调整。

男：很好，你很有厨艺天赋。你喜欢做哪些菜？

女：我喜欢做中国菜，比如宫保鸡丁、糖醋排骨、麻婆豆腐等。

问题：

女士喜欢做哪些菜？

- A. 西餐
- B. 中餐
- C. 日韩料理
- D. 东南亚菜

答案：B

女士喜欢按照食谱做饭，但她也会根据自己的喜好进行一些调整。她掌握了一些基本的烹饪技巧，并具有创新精神。

#### Material 5: De acuerdo con

##### Diálogo

Hombre: Hola, ¿sabes cocinar siguiendo una receta?

Mujer: Sé algunas cosas básicas de cocina. Puedo cocinar según los pasos de una receta, pero a veces hago algunos ajustes para adaptarla a mis propias preferencias.

Hombre: Eso está muy bien, tienes mucho talento para la cocina. ¿Qué tipo de platos te gusta cocinar?

Mujer: Me gustan los platos chinos, como el pollo Gongbao, las costillas agrídulces y el tofu picante.

Pregunta: ¿Qué tipo de platos le gusta cocinar a la señora?

A) Comida occidental B) comida china C) Cocina japonesa y coreana D) Cocina del sudeste asiático

Respuesta: B

A la señora le gusta cocinar siguiendo recetas, pero también hace algunos ajustes para adaptarlas a sus propias preferencias. Domina algunas técnicas básicas de cocina y es innovadora.

En primer lugar, analizamos el contenido de los textos.

- **Tipo de texto:** [4. Excelente] genera con precisión el formato requerido: un diálogo entre un hombre y una mujer, se incluyen todas las 5 palabras proporcionadas en la instrucción. La forma es clara.

- **Longitud del texto:** [1. Insuficiente] La longitud de las oraciones del texto es excesiva, la mayoría de los materiales tienen más de 100 caracteres y el primer material solo tiene 3 oraciones de diálogo.

- **Adecuación al nivel indicado:** [1. Insuficiente] La gramática está adecuada al nivel HSK 4, como en el uso de la oración compuesta condicional “*zhiyao...jiu...* (只要……就……) [siempre y cuando]”. Pero la mayor parte del vocabulario es mucho más difícil que el del nivel 4 del HSK, como por ejemplo “*peixun* (培训) [capacitar]” “*buzhou* (步骤) [paso]” “*miaozhun* (瞄准) [apuntar]” “*weihu* (维护) [mantener]” “*pengren* (烹饪) [cocinar]” “*baoxianshuan* (保险栓) [cerradura]” y el modismo “*honghong-lielie* (轰轰烈烈) [sensacional]”

- **Corrección a nivel gramatical y ortográfico:** [3. Bueno] La mayoría de las oraciones son gramaticalmente correctas, pero en el material 4, la frase “*Ni jide anshi canjia ma?* (你记得按时参加吗?) [¿Recordaste asistir a tiempo?]” tiene un error gramatical y debería cambiarse por “*Ni jide anshi canjia.* (你记得按时参加。)[Recuerda asistir a tiempo.]”

- **Naturalidad de la lengua y coherencia lógica del mensaje:** [2. Aceptable] El lenguaje es algo forzado. Los materiales 1, 2, 3 y 5 empiezan todos con “*nihao* (你好) [hola]”, pero el contenido de la conversación da la impresión de que las dos personas tienen una relación estrecha o están familiarizadas entre sí, y el uso de “*nihao* (你好) [hola]” es un poco oxidado y rígido. Se puede utilizar “*hi* (嗨) [hola]” o “*hei* (嘿) [oye]” para expresar un saludo más íntimo. En particular, en el Material 2, la mujer quiere ir de viaje con el hombre, pero la mujer dice “*nihao* (你好) [hola]” al principio de la conversación y utiliza palabras de cortesía como “*qingwen* (请问) [disculpe]”, que no es un lenguaje muy natural.

- **Autenticidad del texto:** [3. Bueno] La mayoría del texto es auténtico, se ajusta a los temas comunes del examen HSK y refleja la vida real, como amor, viajes, comida, trabajo, pero el material 3 sobre extintores es demasiado complejo y especializado.

En segundo lugar, analizamos los ejercicios diseñados por Gemini.

- **Tipo de ejercicio:** [3. Bueno] Genera correctamente el tipo de pregunta requerido según las instrucciones, y cada pregunta se diseñó con cuatro opciones, pero las respuestas a cada pregunta no se facilitaron al principio.

- **Relación al texto:** [4. Excelente] Las preguntas están estrechamente relacionadas con el texto, bien diseñadas y contribuyen a evaluar la comprensión del contenido.

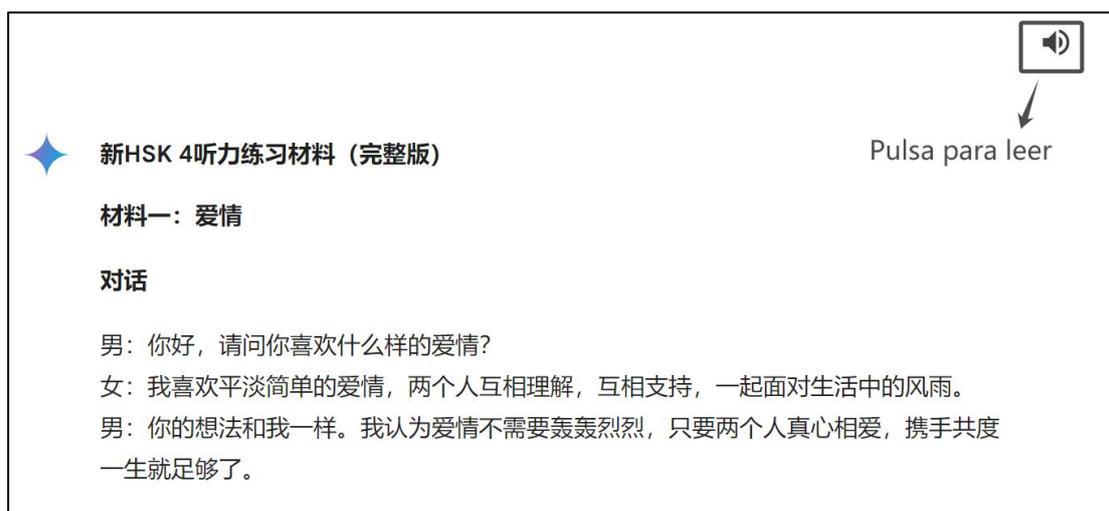
- **Adecuación al nivel indicado:** [3. Bueno] Las preguntas son generalmente comprensibles y solo ocasionalmente contienen palabras nuevas como “*langman-weimei* (浪漫唯美) [romántico y estético]”.

- **Diseño de las preguntas:** [4. Excelente] La mayoría de las preguntas están relacionadas con la segunda mitad del material de comprensión oral, lo que ayuda a comprobar si los alumnos tienen la capacidad de comprender el contenido general del texto y captar la información clave.

- **Diseño de las opciones y respuestas:** [3. Bueno] El diseño de las opciones es muy bueno y flexible. En el Material 1, no sólo se elige la palabra “*honghong-lielie* (轰轰烈烈) [sensacional]” mencionada en el texto como distractor de la palabra “*pingdan-jian dan* (平淡简单) [simple y sencillo]”, sino que la longitud y la estructura de las cuatro opciones son coherentes y pertinentes. En el material 4, además de la opción correcta A), las otras tres opciones también están relacionadas con el tema de la reunión, y el contexto es verdadero, sin errores gramaticales. Sin embargo, el material 3 no sólo tiene una única respuesta correcta. Además de la opción C), la opción B) también es correcta.

#### 5.2.4 Análisis de la dicción oral generado por Gemini

A continuación analizaremos la generación de la dicción oral de Gemini. Pedimos a Gemini que leyera el tercer material de opción múltiple que había generado (Véase Figura 2). A continuación mostraremos los resultados del análisis:



◆ 新HSK 4听力练习材料 (完整版)

材料一: 爱情

对话

男: 你好, 请问你喜欢什么样的爱情?

女: 我喜欢平淡简单的爱情, 两个人互相理解, 互相支持, 一起面对生活中的风雨。

男: 你的想法和我一样。我认为爱情不需要轰轰烈烈, 只要两个人真心相爱, 携手共度一生就足够了。

Pulsa para leer

**Figura 2. Función de lectura de texto de Gemini**

- **Integridad del contenido:** [4. Excelente] Lee el contenido original palabra por palabra, omitir ni añadir nada.

- **Precisión de la pronunciación:** [4. Excelente] La pronunciación es precisa y se ajusta al del mandarín estándar, incluyendo una correcta articulación de las consonantes iniciales y finales, las vocales y los tonos.

- **Velocidad de habla:** [2. Aceptable] Según la grabación de comprensión oral del HSK nivel 4 proporcionada por Zhongwen Kaoshi Fuwu Wang, son aproximadamente 3 caracteres por segundo, y el número de caracteres leídos por Gemini por segundo es 4. Gemini es un poco más rápido y no se puede ajustar manualmente.

- **Cambio de tonos especiales:** [4. Excelente] En el material, la pronunciación de “*bucuo* (不错) [bueno]” es correcta. Como no hay casos de “*yi* (一) [uno]” con el cuarto tono en los materiales, y no hay palabras con tonos neutros, verificamos la pronunciación introduciendo nuevos comandos en Gemini y pidiéndole que leyera en voz alta las respuestas (Véase Extracto 8). Comprobamos que la pronunciación es acertada.

**Extracto 8. Dos palabras con un cuarto tono después del carácter “*yi* (一) [uno]” y dos palabras con tono neutro generadas por Gemini**

**PROMPT:**

请列举出 2 个 “一” 后面跟的是第四声字的词语。

Enumera 2 palabras en las que el carácter “*yi* (一) [uno]” vaya seguida de otra en cuarto tono.

**Gemini:**

好的，以下是两个“一”后面跟的是第四声字的词语：

1. 一次 (*yī cì*): 这个词的意思是“一次”或“一回”。后面的第四声字“次” (*cì*) 表示“时间”或“场合”。
2. 一刻 (*yī kè*): 这个词的意思是“一刻”或“瞬间”。后面的第四声字“刻” (*kè*) 表示“一刻钟”或“短暂的时间”。

Bien, aquí hay dos palabras con un “*yi* (一) [uno]” seguido de un cuarto tono:

1. *yī cì* (一次) [una vez]: Esta palabra significa “una vez”. El carácter con cuarto tono “次” (*cì*) al final significa “tiempo” u “ocasión”.
2. *yī kè* (一刻) [un momento]: Esta palabra significa “un momento” o “un instante”. El carácter con el cuarto tono “刻” (*kè*) al final significa “un cuarto de hora” o “un corto periodo de tiempo”.

**PROMPT:**

请你列举 2 个第二个字读轻声的词语。

Enumera 2 palabras en las que la segunda se pronuncie con tono neutro.

**Gemini:**

好的，以下是两个第二个字读轻声的词语：

1. 妈妈 (mā ma)
2. 爸爸 (bà ba)

Bien, aquí hay dos palabras en las que la segunda se pronuncia con tono neutro:

1. *mama* (妈妈) [madre]
2. *baba* (爸爸) [padre]

- **Entonación:** [4. Excelente] Puede transmitir con precisión las emociones y el tono del texto original como hacer preguntas y dar énfasis.

- **Naturalidad de habla:** [2. Aceptable] A pesar de la precisión y el nivel de pronunciación, el lenguaje en sí es un poco rígido, sobre todo cuando se leen opciones sin pausa, sino de un tirón, la habla es mecanizada y poco natural.

- **Calidad de la voz:** [2. Aceptable] El sonido es claro y melodioso, sin ruidos ni interrupciones, pero no se puede elegir la voz, sólo hay una voz femenina.

- **Facilidad de la operación:** [2. Aceptable] Se puede pausar y reanudar la reproducción en cualquier momento, pero el audio no se puede exportar, no se puede ajustar la velocidad de reproducción y sólo hay una voz femenina para el diálogo hombre-mujer. Es más adecuado para materiales de comprensión oral que sean monólogos independientes.

### 5.3 Comparación entre ChatGPT y Gemini

En la primera y segunda sección, analizamos la eficiencia de los tres ejercicios de comprensión oral del nivel nuevo HSK 4 generados por ChatGPT 3.5 y Gemini, así como sus funciones de conversión de texto a voz. En esta sección, calcularemos el puntaje promedio de la rúbrica para los tres materiales auditivos generados por cada software y compararemos los resultados de ambos. Para una presentación más clara de los resultados, utilizaremos gráficos. Compararemos por separado los resultados de la evaluación del texto, los ejercicios y la dicción oral.

### 5.3.1 Los textos

En primer lugar, presentaremos una tabla con los resultados de la evaluación de textos de ChatGPT 3.5 y Gemini y resumiremos sus puntos fuertes y débiles, y después compararemos los resultados de ambos en un gráfico con explicaciones textuales y conclusiones.

El primero es el resumen de los resultados de la evaluación de los tres materiales de comprensión oral en ChatGPT 3.5.

<b>Resultado de la rúbrica para evaluar los textos de la comprensión oral generados por ChatGPT 3.5</b>				
<b>Criterios</b>	<b>Resultado de cada material</b>			<b>Resultado promedio</b>
	<b>M1</b>	<b>M2</b>	<b>M3</b>	
<b>Tipo de texto</b>	4	4	3	3.67
<b>Longitud del texto</b>	3	4	4	3.67
<b>Adecuación al nivel indicado</b>	1	1	1	1.00
<b>Corrección a nivel gramatical y ortográfico</b>	3	4	3	3.33
<b>Naturalidad de la lengua y coherencia lógica del mensaje</b>	2	1	3	2.00
<b>Autenticidad del texto</b>	4	4	4	4.00
<b>Conclusión</b>	Los materiales muestran fortalezas en términos de autenticidad, longitud y tipo de texto. Sin embargo, hay debilidades significativas en la adecuación al nivel indicado y en la naturalidad y coherencia de la lengua.			

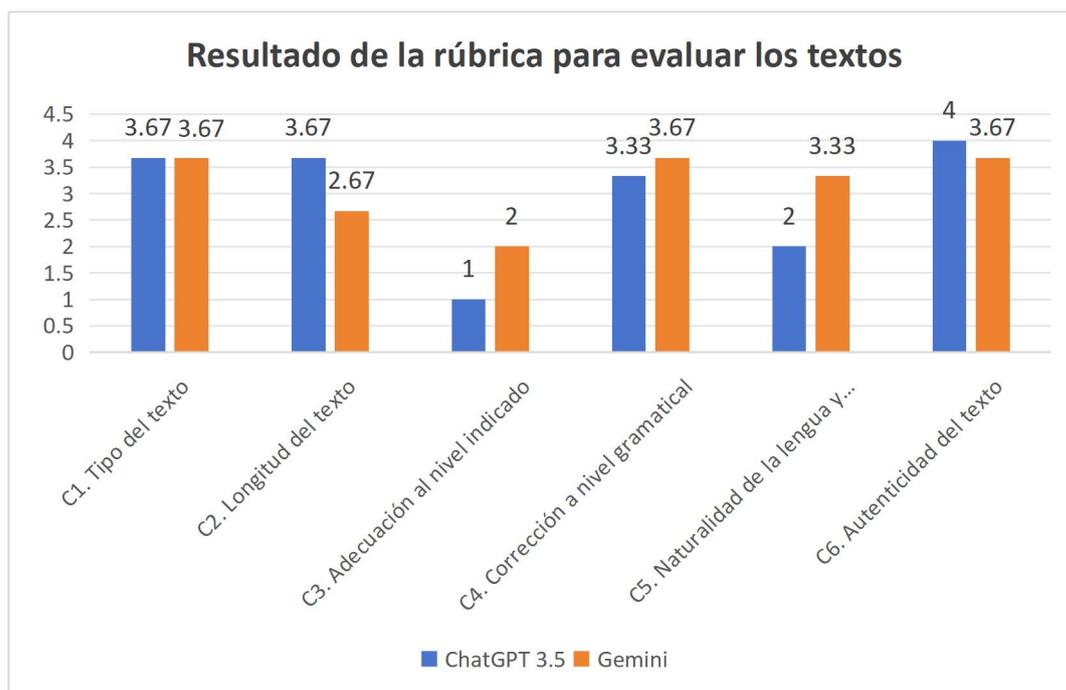
**Tabla 9. Resultado de la rúbrica para evaluar los textos de la comprensión oral generados por ChatGPT 3.5**

El segundo es el resumen de los resultados de la evaluación de los tres materiales de comprensión oral en Gemini.

<b>Resultado de la rúbrica para evaluar los textos de la comprensión oral generados por Gemini</b>				
<b>Criterios</b>	<b>Resultado de cada material</b>			<b>Resultado promedio</b>
	<b>M1</b>	<b>M2</b>	<b>M3</b>	
<b>Tipo de texto</b>	<b>3</b>	<b>4</b>	<b>4</b>	<b>3.67</b>
<b>Longitud del texto</b>	<b>3</b>	<b>4</b>	<b>1</b>	<b>2.67</b>
<b>Adecuación al nivel indicado</b>	<b>2</b>	<b>3</b>	<b>1</b>	<b>2.00</b>
<b>Corrección a nivel gramatical y ortográfico</b>	<b>4</b>	<b>4</b>	<b>3</b>	<b>3.67</b>
<b>Naturalidad de la lengua y coherencia lógica del mensaje</b>	<b>4</b>	<b>4</b>	<b>2</b>	<b>3.33</b>
<b>Autenticidad del texto</b>	<b>4</b>	<b>4</b>	<b>3</b>	<b>3.67</b>
<b>Conclusión</b>	Los materiales muestran fortalezas en términos de tipo de texto, corrección gramatical, naturalidad y coherencia, y autenticidad. Sin embargo, hay debilidades significativas en la longitud del texto y en la adecuación al nivel indicado.			

**Tabla 10. Resultado de la rúbrica para evaluar los textos de la comprensión oral generados por Gemini**

Ahora utilizaremos un gráfico para mostrar los resultados de las puntuaciones de los dos softwares de Inteligencia Artificial.



**Figura 3. Comparación del resultado de la rúbrica para evaluar los textos generados por ChatGPT 3.5 y Gemini**

Como podemos visualizar en el gráfico, Gemini supera a ChatGPT 3.5 en la mayoría de los criterios, excepto en la imprecisión de la longitud del texto y en la autenticidad del texto, y, en particular, obtiene mejores resultados en cuanto a precisión gramatical, naturalidad y coherencia lingüística. Ambas herramientas fueron capaces de comprender bien las instrucciones y, en última instancia, generar tipos de texto que cumplieran los requisitos, incorporaban las palabras dadas en las instrucciones y se presentaban con claridad, con temas materiales auténticos y cercanos a la vida real. Sin embargo, ninguno de los dos programas fue capaz de generar contenidos del nivel nuevo HSK 4 especificado, y los contenidos generados eran demasiado sencillos o demasiado difíciles.

### 5.3.2 Los ejercicios

Antes que nada, mostraremos el resultado de la rúbrica para evaluar los ejercicios de la comprensión oral generados por ChatGPT 3.5.

<b>Resultado de la rúbrica para evaluar los ejercicios de la comprensión oral generados por ChatGPT 3.5</b>				
<b>Criterios</b>	<b>Resultado de cada material</b>			<b>Resultado promedio</b>
	<b>M1</b>	<b>M2</b>	<b>M3</b>	
<b>Tipo de ejercicio</b>	4	4	4	<b>4.00</b>
<b>Relación al texto</b>	3	4	4	<b>3.67</b>
<b>Adecuación al nivel indicado</b>	4	4	3	<b>3.67</b>
<b>Diseño de las preguntas</b>	2	2	4	<b>2.67</b>
<b>Diseño de las opciones y respuestas</b>	2	3	2	<b>2.33</b>
<b>Conclusión</b>	Los ejercicios presentan fortalezas significativas en términos de tipo, relación con el texto y adecuación al nivel. Sin embargo, hay áreas críticas que requieren mejora en el diseño de las preguntas y las opciones de respuesta.			

**Tabla 11. Resultado de la rúbrica para evaluar los ejercicios de la comprensión oral generados por ChatGPT 3.5**

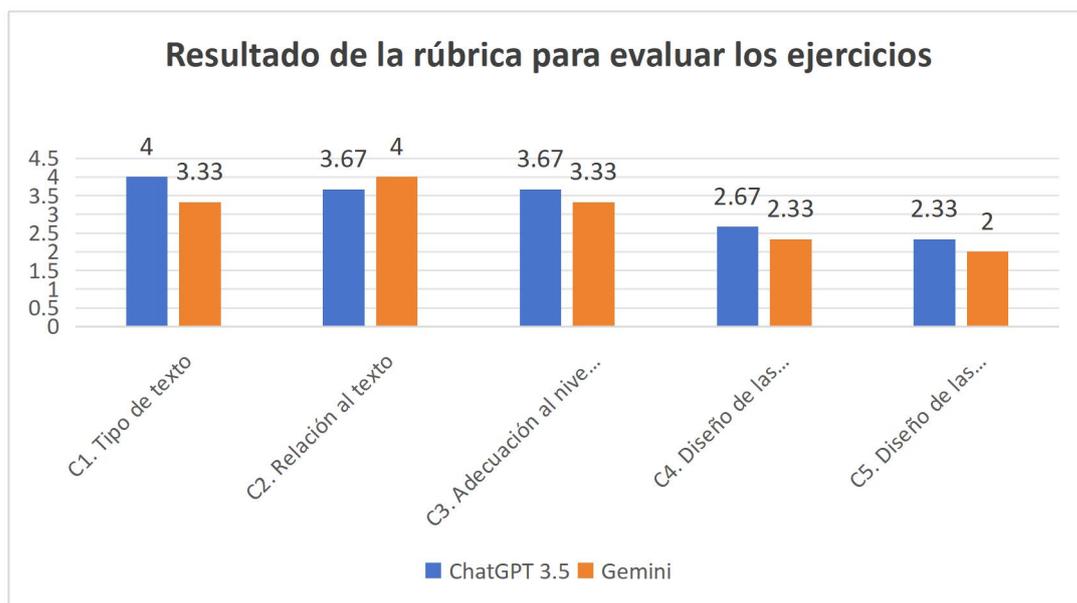
Consecutivamente, mostraremos el resultado de la rúbrica para evaluar los ejercicios de la comprensión oral generados por Gemini.

<b>Resultado de la rúbrica para evaluar los ejercicios de la comprensión oral generados por Gemini</b>				
<b>Criterios</b>	<b>Resultado de cada material</b>			<b>Resultado promedio</b>
	<b>M1</b>	<b>M2</b>	<b>M3</b>	
<b>Tipo de ejercicio</b>	4	3	3	<b>3.33</b>
<b>Relación al texto</b>	4	4	4	<b>4.00</b>
<b>Adecuación al nivel indicado</b>	4	3	3	<b>3.33</b>
<b>Diseño de las preguntas</b>	1	2	4	<b>2.33</b>
<b>Diseño de las opciones y respuestas</b>	1	2	3	<b>2.00</b>
<b>Conclusión</b>	Los ejercicios muestran fortalezas en la			

relación con el texto y son razonablemente adecuados en el tipo de ejercicio y nivel. Sin embargo, hay áreas críticas que requieren mejora en el diseño de las preguntas y las opciones de respuesta.

**Tabla 12. Resultado de la rúbrica para evaluar los ejercicios de la comprensión oral generados por Gemini**

A continuación, vamos a usar un gráfico para mostrar las puntuaciones obtenidas por ambos programas de inteligencia artificial.



**Figura 4. Comparación del resultado de la rúbrica para evaluar los ejercicios generados por ChatGPT 3.5 y Gemini**

El gráfico muestra que, en lo que respecta al diseño de las preguntas de comprensión oral, a excepción del segundo punto, la relevancia de las preguntas con respecto al texto original, en el que ChatGPT 3.5 obtiene el resultado inferior, ChatGPT 3.5 supera a Gemini en todos los demás aspectos. ChatGPT 3.5 genera el tipo correcto de preguntas de acuerdo con las instrucciones, tiene mejor capacidad de imitación y presenta el contenido de forma clara y concisa, mientras que Gemini no es tan claro y preciso en la presentación del contenido y es más débil en capacidad de imitación.

Ambos tenían fallos en el diseño de las preguntas y las opciones, pero ChatGPT 3.5 era a veces más flexible a la hora de diseñar preguntas con ideas ingeniosas que requerían que los alumnos hicieran más inferencias o escucharan atentamente los detalles. Gemini, por su parte, es más rígido, diseñando preguntas sencillas o que pueden responderse utilizando el sentido común. Sin embargo, ambos programas de IA proporcionaban a veces respuestas incorrectas o respuestas inventadas que requerían una revisión posterior por parte del docente.

### 5.3.3 La dicción oral

Para empezar, presentaremos el resultado de la rúbrica para evaluar la dicción oral de la comprensión oral generados por ChatGPT 3.5.

<b>Resultado de la rúbrica para evaluar la dicción oral de la comprensión oral generados por ChatGPT 3.5</b>	
<b>Criterios</b>	<b>Resultado</b>
<b>Integridad del contenido</b>	<b>1</b>
<b>Precisión de pronunciación</b>	<b>3</b>
<b>Velocidad de habla</b>	<b>1</b>
<b>Cambio de tonos especiales</b>	<b>4</b>
<b>Entonación</b>	<b>4</b>
<b>Naturalidad de habla</b>	<b>4</b>
<b>Calidad de la voz</b>	<b>4</b>
<b>Facilidad de la operación</b>	<b>1</b>
<b>Nota final/ Puntuación total</b>	<b>22/32</b>
<b>Conclusión</b>	El desempeño fue sobresaliente en el manejo del tono y la naturalidad de habla, así como la calidad de la voz. Sin embargo, hay áreas significativas de mejora en la integridad del

	contenido, velocidad de habla y facilidad de operación.
--	---

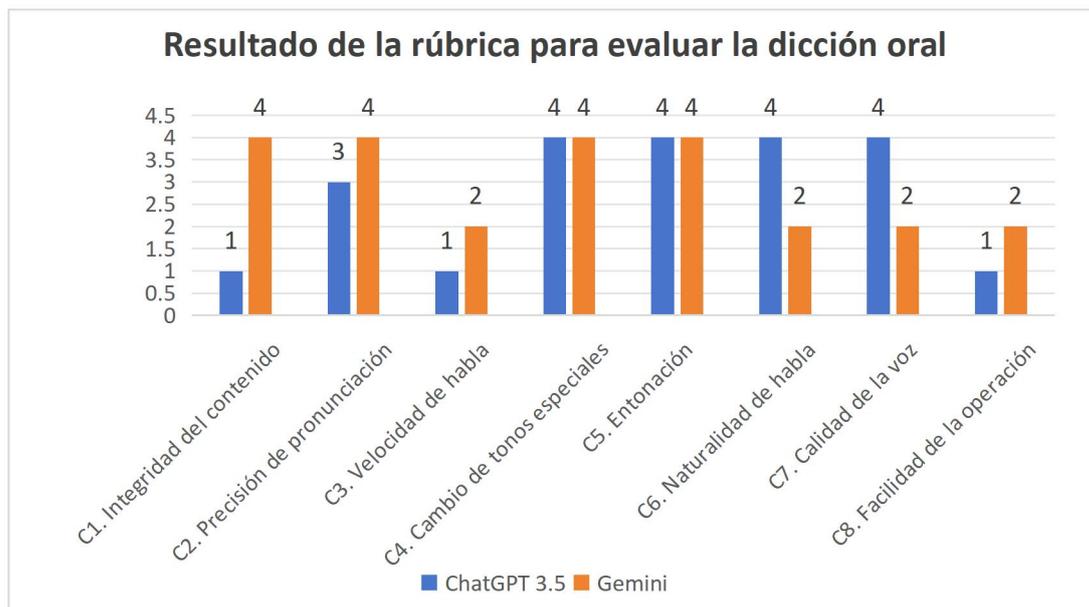
**Tabla 13. Resultado de la rúbrica para evaluar la dicción oral de la comprensión oral generados por ChatGPT 3.5**

Seguidamente, mostraremos el resultado de la rúbrica para evaluar la dicción oral de la comprensión oral generados por Gemini.

<b>Resultado de la rúbrica para evaluar la dicción oral de la comprensión oral generados por Gemini</b>	
<b>Criterios</b>	<b>Resultado</b>
<b>Integridad del contenido</b>	<b>4</b>
<b>Precisión de pronunciación</b>	<b>4</b>
<b>Velocidad de habla</b>	<b>2</b>
<b>Cambio de tonos especiales</b>	<b>4</b>
<b>Entonación</b>	<b>4</b>
<b>Naturalidad de habla</b>	<b>2</b>
<b>Calidad de la voz</b>	<b>2</b>
<b>Facilidad de la operación</b>	<b>2</b>
<b>Nota final/ Puntuación total</b>	<b>24/32</b>
<b>Conclusión</b>	El desempeño fue sobresaliente en la integridad del contenido, la precisión de pronunciación y el tono de habla, incluidos los tonos especiales. Pero hay mucho margen de mejora en cuanto a la velocidad del habla, la naturalidad, la calidad de la voz y facilidad de la operación.

**Tabla 14. Resultado de la rúbrica para evaluar la dicción oral de la comprensión oral generados por Gemini**

A partir de los resultados de las respectivas evaluaciones, se utilizará un icono para visualizarlas y compararlas.



**Figura 5. Comparación del resultado de la rúbrica para evaluar la dicción oral generada por ChatGPT 3.5 y Gemini**

En los resultados de la dicción oral, aunque Gemini tiene una puntuación global más alta que ChatGPT 3.5, hay que señalar que ChatGPT 3.5 y Gemini tienen sus propios puntos fuertes y débiles. Gemini es mejor en la integridad de la lectura del contenido y la precisión de la pronunciación, mientras que ChatGPT 3.5 tiene el problema de omitir o no poder leer correctamente el material en chino. En cuanto a la naturalidad del habla, la fluidez y la diversidad de voces, ChatGPT 3.5 es más fluido y natural y ofrece una variedad de voces masculinas y femeninas, mientras que Gemini es más mecanizado y sólo tiene una voz femenina. Cabe destacar que ambos programas transmiten correctamente la emoción y el tono del texto original, como el tono interrogativo o la indicación de énfasis, y sobre todo los tonos especiales también se pronuncian correctamente. Sin embargo, la velocidad de la voz en ambas herramientas es relativamente rápida y no se puede ajustar, y ambos materiales solamente permiten una voz durante la lectura, es decir, no se pueden usar voces distintas en el caso de diálogos. Además, ChatGPT 3.5 no permite hacer pausas durante la reproducción, lo que obliga a empezar de nuevo la lectura cada vez que se

reproduce, lo que hace más engorroso su manejo para el docente si se va a reproducir directamente como audio del material de comprensión oral.

## **6. ANÁLISIS DEL PROCESO**

En la sección de metodología, mencionamos que esta tesina tiene dos objetivos específicos, uno es evaluar el producto final de comprensión oral creados por herramientas de IA, el otro es determinar en qué medida es eficiente el uso de herramientas de IA como asistente del docente de chino, es decir, examinar el proceso hasta conseguir los ejercicios. Porque si se necesitan más pasos para conseguir el resultado que queremos, no se reduce el trabajo del profesor, sino que se pierde más tiempo.

Así pues, en esta sección hablaremos primero de la eficacia de las dos herramientas a la hora de comunicarse entre sí, de los problemas encontrados y de las soluciones finales, y después contaremos el número de pasos que tardan ChatGPT 3.5 y Gemini en generar 3 materiales de comprensión oral. Por último, utilizaremos un gráfico para hacer una comparación visual.

### **6.1 Análisis de ChatGPT**

El primero es ChatGPT 3.5, afortunadamente, el uso de ChatGPT 3.5 es muy ágil, después de determinar las instrucciones, ChatGPT 3.5 a menudo puede generar a la vez el contenido y el formato que queremos (Véase Figura 6), en el que imita bien las preguntas de ejemplo, el número de serie del material, las preguntas, las respuestas se presentan claramente, no hay información superflua, el material puede ser utilizado directamente para analizar.

听力材料：现在我很少看电视，其中一个原因是，广告太多了，不管什么时间，也不管什么节目，只要你打开电视，总能看到那么多的广告，浪费我的时间。题目：他喜欢看电视广告。答案：x。这是一道新HSK四级考试的听力材料和题目，请你模仿它，生成5道听力题，听力材料30到60字，材料中分别包含以下词汇：爱情、安排、安全、按时、按照。题目答案有对有错。

🗣️ 听力材料1：我认为爱情是生活中美好的一部分，但是对我来说，最重要的是安排好我的工作和生活。  
 题目：他觉得爱情很重要。  
 答案：√

听力材料2：在旅行中，我们要时刻注意安全，尤其是在陌生的地方，不要随便相信陌生人。  
 题目：在旅行中不需要注意安全。  
 答案：x

听力材料3：我总是按时完成工作，因为我知道时间很宝贵，不能浪费。  
 题目：他经常迟到。  
 答案：x

听力材料4：按照老师的安排，我们这个学期要完成一份大作业，需要分组合作。  
 题目：他们没有大作业要完成。  
 答案：x

听力材料5：我按照自己的计划安排了一次浪漫的约会，但是天公不作美，下着大雨。  
 题目：他的约会很顺利。  
 答案：x

**Figura 6. El material de comprensión oral generado por ChatGPT 3.5**

Por lo tanto, ChatGPT 3.5 requiere 0 pasos después de introducir el comando y hasta que se genere el material que queremos, no se necesitan más comandos extra.

## 6.2 Análisis de Gemini

Sin embargo, el uso de Gemini no es tan fluido. Después de introducir las instrucciones, para los materiales generados por Gemini, a menudo tenemos que repetir o introducir nuevas instrucciones para que realice el siguiente paso de optimización, o incluso varios pasos de optimización, para conseguir el texto final que deseamos. A continuación, explicaremos el proceso de generación de los 3 materiales respectivamente, así como los problemas encontrados y las soluciones tomadas.

Gemini generó inicialmente contenidos del material 1 en inglés (Véase Figura 7), que no se ajustaban al formato que queríamos. Por lo tanto, volvimos a pedirle que tradujera del inglés al chino. (Véase Figura 8)

听力材料：现在我很少看电视，其中一个原因是，广告太多了，不管什么时间，也不管什么节目，只要你打开电视，总能看到那么多的广告，浪费我的时间。题目：他喜欢看电视广告。答案：×。这是一道新HSK四级考试的听力材料和题目，请你模仿它，生成5道听力题，听力材料30到60字，材料中分别包含以下词汇：爱情、安排、安全、按时、按照。题目答案有对有错。

◆ Sure, here are 5 HSK 4 Listening Practice Questions that follow the same format as the original example:

**Question 1:**

**Listening Material:**

小丽很喜欢看电影，她经常周末去电影院看电影。最近，她听说了一部新电影，非常想去看，于是她就提前在网上订好了电影票。

**Question:** 小丽不喜欢看电影。 (Yes/No)

**Answer:** ×

**Question 2:**

**Listening Material:**

王先生是一位非常注重安全的人，他家里的所有电器都安装了漏电保护装置，他还经常

→ El inglés

**Figura 7. El material con el inglés y el chino generado por Gemini**

Listening Material:  
张老师是一位经验丰富的老师，她总是按照教学计划上课，从不随意更改课程内容。她也经常给学生布置作业，并认真批改作业。

Question: 张老师不按照教学计划上课。 (Yes/No)

Answer: × 把这里的所有英文都翻译成中文。

Pedimos a Gemini que traduzca del inglés al chino.

◆ Sure, here are the questions and answers translated into Chinese:

**题目1:**

**听力材料:**

小丽很喜欢看电影，她经常周末去电影院看电影。最近，她听说了一部新电影，非常想去看，于是她就提前在网上订好了电影票。

**问题:** 小丽不喜欢看电影。 (是/否)

**答案:** 否

**解释:**

文本中明确指出小丽很喜欢看电影，经常周末去电影院看电影，并提前在网上订好了新电影的票，因此答案为“否”。

显示草稿

**Figura 8. Gemini traduce el material del inglés al chino**

Al generar el material 2, también nos encontramos con una situación en la que aparecía la lengua inglesa, y se nos dio de nuevo la instrucción de pedir a Gemini que

contestara sólo en chino. Por eso se sugiere que los profesores añadan a la instrucción “Por favor, conteste sólo en chino” al principio.

El proceso de generación del material 3 fue algo complicado. En primer lugar, tras introducir las instrucciones, volvimos a encontrarnos con el mismo problema que en los dos materiales anteriores, es decir, aparecía el inglés y, además, Gemini establecía dos preguntas de opción múltiple por material, cada una de ellas con sólo tres opciones, lo que no coincide con los ejemplos que proporcionamos (Véase Figura 9).

◆ New HSK 4 Listening Questions

Material 1: Love

Dialogue:

Man: 你好, 请问你喜欢什么样的爱情?  
Woman: 我喜欢平淡简单的爱情, 两个人互相理解, 互相支持, 一起面对生活中的风雨。  
Man: 你的想法和我一样。我认为爱情不需要轰轰烈烈, 只要两个人真心相爱, 携手共度一生就足够了。

Questions:

1. 女士喜欢什么样的爱情?  
A. 轰轰烈烈的爱情  
B. 平淡简单的爱情  
C. 充满挑战的爱情
2. 男士对爱情有什么看法?  
A. 他认为爱情应该轰轰烈烈。  
B. 他认为爱情应该平淡简单。  
C. 他认为爱情应该充满挑战。

Answer:

**Figura 9. El material mal generado por Gemini**

Así que volvimos a introducir el comando para que se modificara, pero el resultado mostró que Gemini sólo había reducido 2 preguntas a 1, y los otros problemas todavía se mantenían. (Véase Figura 10).

1. 纯中文回答 2. 每段材料仅设置一道选择题，每道选择题有A、B、C、D四个选项。

◆ New HSK 4 Listening Questions (Pure Chinese Response)

**Material 1: Love**

**Dialogue:**

**Man:** 你好，请问你喜欢什么样的爱情？

**Woman:** 我喜欢平淡简单的爱情，两个人互相理解，互相支持，一起面对生活中的风雨。

**Man:** 你的想法和我一样。我认为爱情不需要轰轰烈烈，只要两个人真心相爱，携手共度一生就足够了。

**Question:**

女士认为爱情应该是什么样的？

A. 轰轰烈烈  
B. 平淡简单  
C. 充满挑战

**Figura 10. Gemini no puede modificar bien a las instrucciones**

Por lo tanto, creamos una nueva ventana de chat con Gemini, copiamos los resultados originales, volvimos a introducir los comandos y, en lugar de pedir que se redujera el número de preguntas a una, sólo pedimos que se tradujera al chino y aumentara el número de opciones porque nos preocupaba que Gemini no fuera capaz de manejar más de un comando a la vez.

Los resultados mostraron que Gemini sólo traducía al chino la parte en inglés del texto original, sin cambiar el número de opciones (Véase Figura 11), y al final del material, Gemini intentaba empezar a traducir de nuevo todo el texto al inglés. Tuvimos que introducir un comando para finalizar la traducción y pedir a Gemini que aumentara el número de opciones a cuatro (Véase Figura 12).

请将上述这段话中的英文翻译成纯中文，并且每道题目设计4个选项，目前只有3个。

🔊

◆ **新HSK 4听力练习材料**

**材料一：爱情**

对话：

男：你好，请问你喜欢什么样的爱情？  
 女：我喜欢平淡简单的爱情，两个人互相理解，互相支持，一起面对生活中的风雨。  
 男：你的想法和我一样。我认为爱情不需要轰轰烈烈，只要两个人真心相爱，携手共度一生就足够了。

问题：

1. 女士喜欢什么样的爱情？

- A. 轰轰烈烈的爱情
- B. 平淡简单的爱情
- C. 充满挑战的爱情

2. 男士对爱情有什么看法？

- A. 他认为爱情应该轰轰烈烈。
- B. 他认为爱情应该平淡简单。
- C. 他认为爱情应该充满挑战。

**Figura 11. Gemini no puede modificar bien a las instrucciones**

🐾 不需要英文翻译，每道题设置四个选项。（A.B.C.D）

🔊

◆ **新HSK 4听力练习材料（完整版）**

**材料一：爱情**

对话：

男：你好，请问你喜欢什么样的爱情？  
 女：我喜欢平淡简单的爱情，两个人互相理解，互相支持，一起面对生活中的风雨。  
 男：你的想法和我一样。我认为爱情不需要轰轰烈烈，只要两个人真心相爱，携手共度一生就足够了。

问题：

1. 女士喜欢什么样的爱情？

- A. 轰轰烈烈的爱情
- B. 平淡简单的爱情
- C. 充满挑战的爱情
- D. 浪漫唯美的爱情

**Figura 12. Ingresamos nuevamente el comando para que Gemini lo modifique**

Tras introducir sólo una petición cada vez, Gemini aumentó el número de opciones a cuatro. Inmediatamente después, introdujimos de nuevo la orden para pedir a Gemini que estableciera sólo una pregunta de opción múltiple para cada material, y finalmente conseguimos lo que queríamos. En este punto, sin embargo, nos dimos cuenta de que habíamos omitido la respuesta a la pregunta, por lo que volvimos a introducir el comando para pedirle que añadiera la respuesta (Véase Figura 13). Por último, integramos las preguntas y respuestas del material para obtener el contenido completo del Material 3 presentado anteriormente.



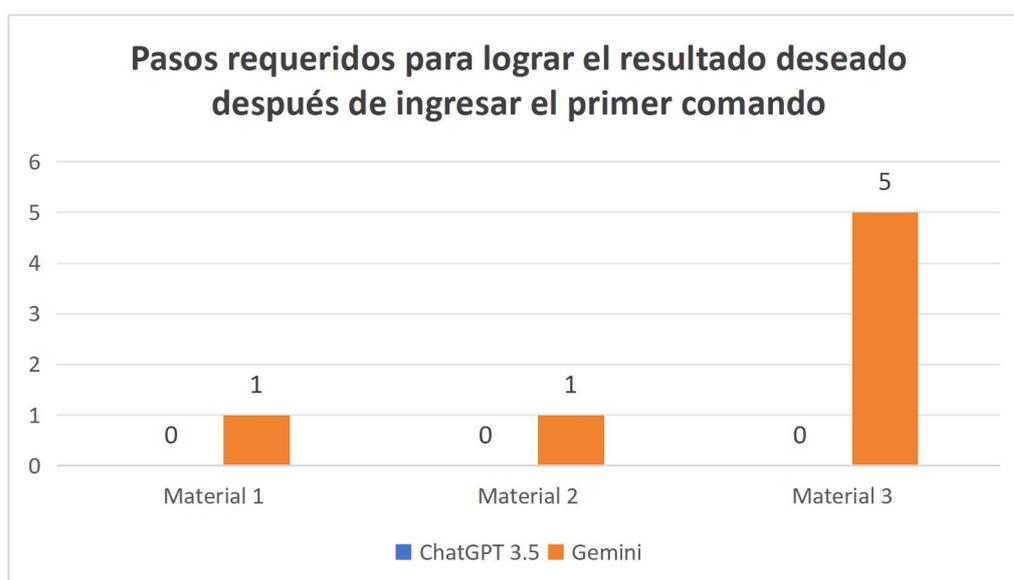
**Figura 13. Pedimos a Gemini que proporcionara respuestas a las preguntas anteriores**

Por lo general, al ingresar el mismo comando que ChatGPT3.5, Gemini no puede generar el contenido que queremos de una sola vez. Al optimizar, no puede modificar todos los problemas existentes de acuerdo con el contenido de las instrucciones. En

lugar de ello, se deben repetir varios pasos y es importante solo modificar un problema cada vez. Desde el momento en que introdujimos los comandos en Gemini hasta el momento en que Gemini generó lo que queríamos, introdujimos los comandos 1 vez extra para los materiales 1 y 2, e introdujimos los comandos 5 veces extra para el material 3.

### 6.3 Comparación entre ChatGPT y Gemini

Del análisis anterior se desprende fácilmente que ChatGPT es más eficiente que Gemini en la generación de los materiales de comprensión oral. También podemos visualizarlo mejor a través del siguiente gráfico.



**Figura 14. Comparación de los pasos requeridos para conseguir el resultado deseado de ChatGPT 3.5 y Gemini**

Cuando introdujimos las mismas instrucciones, ChatGPT entendió bien los requisitos de las instrucciones e imitó bien los elementos de las preguntas de ejemplo, presentando claramente el contenido del material auditivo, las preguntas y proporcionando las respuestas, y a menudo incrustando vocabulario en el material según se requería. Gemini, por otro lado, a menudo piensa más allá de generar el contenido, por ejemplo proporciona explicaciones de las respuestas, o hace

traducciones al inglés, o genera múltiples preguntas como en el Material 3, que no es tan claro como ChatGPT. Al mismo tiempo, Gemini no puede mejorar más de un punto a la vez cuando optimiza, sino sólo un punto cada vez, por lo que se necesitan más pasos para generar el material de comprensión oral final que cumpla nuestros requisitos.

## 7. CONCLUSIONES

Ahora, con respecto a la pregunta de qué herramienta resultará más útil para generar materiales de comprensión oral en clases de CLE, después de analizar los textos, ejercicios y la dicción oral generados por ChatGPT 3.5 y Gemini, podemos llegar a la siguiente conclusión: Gemini y ChatGPT 3.5 tienen sus propias ventajas e inconvenientes a la hora de generar material de comprensión oral.

Los textos generados por Gemini son más naturales, lógicos y correctos a nivel gramatical. Sin embargo, ChatGPT 3.5 es más flexible en el diseño de preguntas y presenta un formato más claro que Gemini. En cuanto a la dicción oral, ChatGPT 3.5 obtiene mejores resultados en la naturalidad del lenguaje y calidad del sonido, mientras que Gemini obtiene mejores resultados en la precisión de la pronunciación y velocidad del habla. Por último, en relación con la eficacia de la interacción con estos LLMs, ChatGPT 3.5 es mucho más eficiente, ya que a menudo comprende la totalidad de las instrucciones de una sola vez y genera material que cumple nuestros requisitos, mientras que Gemini, ante las mismas instrucciones, requiere más pasos para optimizar.

En base a estas conclusiones generales, hemos elaborado las dos sugerencias siguientes para tal situación:

1. El docente puede combinar las ventajas de estos dos programas de IA. Por ejemplo, utilizar Gemini para generar contenidos de texto y ChatGPT para el diseño de los ejercicios. Los alumnos con un nivel bajo de chino pueden utilizar Gemini con una velocidad de habla moderada para practicar la comprensión oral, y los alumnos con un nivel alto de chino pueden utilizar ChatGPT para practicar.
2. En cuanto a las deficiencias del software, el docente también puede resolverlas si sabe cómo utilizarlo de manera eficiente. Por ejemplo, puede que a veces la IA no sea capaz de generar el texto exactamente cuando hay muchas instrucciones, así que, para facilitar la optimización posterior, puede pedirle que genere sólo un ejercicio, o que

introduzca sólo una instrucción cada vez. Generar sólo un ejercicio a la vez también es más útil para que los profesores reproduzcan la grabación repetidamente en clase.

Sin embargo, ChatGPT y Gemini comparten algunos de los mismos problemas: por ejemplo, un defecto recurrente es que no pueden generar textos adecuados al nivel indicado como el nuevo HSK 4. Al mismo tiempo, hay que seguir mejorando la flexibilidad en el diseño de las preguntas, la corrección a nivel gramatical y la precisión de las respuestas. Además, la función de texto a voz no permite ajustar la velocidad del habla ni alternar voces masculinas y femeninas para los diálogos, lo que resulta inconveniente si lo que se busca es imitar el tipo de diálogos que suelen aparecer en los exámenes de lengua.

Por lo tanto, sugerimos que los profesores puedan pedir a ChatGPT o Gemini que incorporen las palabras o los puntos gramaticales que queremos examinar para que el contenido coincida con el nivel del idioma que queremos. Por otro lado, al utilizar el software de IA, el docente no debe confiar completamente en el contenido generado y necesita comprobarlo más a fondo, ya que a veces se generan respuestas erróneas u oraciones que contienen problemas gramaticales. Además, se podrían buscar opciones distintas para complementar las limitaciones en cuanto a la dicción oral. Por ejemplo, se pueden utilizar algunas herramientas de conversión de texto a voz que permiten cambiar la velocidad de habla y elegir voces, como narakeet<sup>9</sup>: en este caso, podemos copiar y pegar el texto generado por la IA en esta herramienta, seleccionar la voz que deseemos (hombre, mujer o niño), establecer la velocidad de habla y luego hacer clic para generar una grabación. Se puede reproducir la grabación en línea o descargarla localmente.

Este estudio también presenta algunas limitaciones. En primer lugar, hemos utilizado la versión gratuita de ambas herramientas, pero durante los experimentos tuvimos algunas oportunidades de usar ChatGPT-4o de forma gratuita. Descubrimos que los

---

<sup>9</sup> <https://www.narakeet.com/app/text-to-audio/>

resultados obtenidos con ChatGPT-4o fueron mejores en comparación con ChatGPT-3.5. Por lo tanto, si se utilizara versiones de pago como ChatGPT-4 o ChatGPT-4o, así como Gemini Advanced, los resultados podrían ser superiores. En segundo lugar, solo hemos solicitado a ambos programas generar materiales de comprensión oral para el nivel nuevo HSK4. Tal vez, al generar materiales para niveles más bajos de chino, como HSK1-3, el nivel de vocabulario y gramática cumpliría mejor con nuestros requisitos.

Durante el trabajo, también descubrimos algunos problemas que podrían proporcionar nuevas direcciones de investigación para futuros investigadores:

Por un lado, la inteligencia artificial generativa podría producir contenido sesgado o con prejuicios de género. En el tercer material de comprensión oral generado por Gemini, específicamente en el material cinco, un hombre le pregunta a una mujer si sabe cocinar siguiendo una receta y comenta sobre sus habilidades culinarias: “*Henhao, ni henyou chuyi tianfu.* (很好, 你很有厨艺天赋。) [Eso está muy bien, tienes mucho talento para la cocina.]”. Este tipo de diálogo reproduce estereotipos de género.

Por otro lado, en esta tesina, hemos utilizado ChatGPT y Gemini para generar materiales de comprensión oral en chino. Sin embargo, debido a la influencia del inglés, la lengua mayoritaria de los textos que conforman estos LLMs, los textos pueden presentar deficiencias en cuanto a naturalidad y precisión lingüística. En investigaciones futuras, se podrían emplear modelos de lenguaje basados en escenarios en chino, como ERNIE Bot (Wenxin Yiyan)<sup>10</sup>, SparkDesk (Xunfei Xinghuo)<sup>11</sup>, y Tongyi Qianwen<sup>12</sup> para evaluar si pueden ofrecer un mejor rendimiento en la generación de materiales didácticos para la enseñanza del chino como lengua extranjera.

---

<sup>10</sup> <https://yiyun.baidu.com/>

<sup>11</sup> <https://xinghuo.xfyun.cn/>

<sup>12</sup> <https://tongyi.aliyun.com/qianwen/>

Con el trabajo hemos aprendido mucho. Primero, al revisar y analizar una gran cantidad de literatura, hemos profundizado en la historia y el estado actual del desarrollo de la inteligencia artificial, que es un campo de conocimiento completamente nuevo para nosotros que ha ampliado nuestra perspectiva. Segundo, durante la investigación, inevitablemente hemos enfrentado dificultades y obstáculos. Esto fortaleció nuestra capacidad para pensar de manera multifacética y nos alentó a buscar la ayuda de nuestra tutora, combinando diversos recursos para resolver problemas. Además, completar esta tesina requirió una gran cantidad de tiempo, un alto grado de concentración y una fuerte capacidad de automotivación. Como resultado, nuestra autodisciplina ha mejorado, lo que también será beneficioso para nuestro desempeño en futuros trabajos.

En definitiva, el software de IA sólo se recomienda como una ayuda para la enseñanza y no como un reemplazo completo del trabajo del docente, y en la etapa inicial de uso de las herramientas de IA, se necesita mucho tiempo para refinar las instrucciones y entender mejor cómo funciona la IA, y si el docente no está familiarizado con estos software, entonces también puede agravar su trabajo. Sin embargo, si a través de un entrenamiento continuo se adquieren buenas habilidades de comunicación y comandos efectivos, recomendamos intentar utilizar la inteligencia artificial. Esto podría ofrecer nuevas ideas y puntos de innovación para los contenidos educativos. Además, al asegurar la calidad del contenido, también puede reducir eficazmente la carga de trabajo de los docentes.

## 8. BIBLIOGRAFÍA

- Ahmed, I., Roy, A., Kajol, M., Hasan, U., Datta, P. P., y Reza, M. R. (2023). ChatGPT vs. Bard: a comparative study. *Authorea Preprints*. 2-10.
- Baskara, R. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2), 343-358.
- Borji, A. (2023). A categorical archive of chatgpt failures. arXiv preprint arXiv:2302.03494.
- Bowman, E. (2022). A new AI chatbot might do your homework for you. But it's still not an A+ student. NPR. Retrieved January, 5.
- Cádiz Deleto, J. L. (2002). Presentación. En Consejo de Europa (ed.), *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Artes Gráficas Fernández Ciudad, S. L. ISBN: 84-667-1618-1.
- Cao Hui (曹慧). (2000). “Tingli yuliao fenxi yu tingli jiaocai de bianxie (听力语料分析与听力教材的编写) [Análisis del corpus de comprensión oral y redacción de materiales didácticos de comprensión oral]”. *Yuyan Wenhua Jiaoxue Yanjiu Jikan (语言文化教学研究集刊) [Revista de investigación sobre enseñanza de lengua y cultura]*, 4, 11-31. Huayu Jiaoxue Chubanshe (华语教学出版社) [Prensa docente china].
- Casas-Tost, Helena, y Rovira-Esteva, Sara. (2018). La adaptación del chino al MCER en España: un análisis crítico. *Inter Asia papers*, (61), 1-36.
- Chaka, C. (2023). Generative AI chatbots-ChatGPT versus YouChat versus Chatsonic: Use cases of selected areas of applied English language studies. *International Journal of Learning, Teaching and Educational Research*, 22(6), 1-19.

- Consejo de Europa. (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Artes Gráficas Fernández Ciudad, S. L. ISBN: 84-667-1618-1.
- Dong Xiaodai (董晓岱) y Xie Wei (谢卫). (2012). “Duiwai hanyu tingli jiaoxue chutan (对外汉语听力教学初探) [Un estudio preliminar sobre la enseñanza de la escucha del chino como lengua extranjera]”. *Sanwen Baijia (Xin Yuwen Huoye) [散文百家(新语文活页)] [Cientos de escuelas de prosa (nuevo libro chino de hojas sueltas)]*, (3).
- Du Shisheng (杜氏盛). (2013). “Qiantan yuenan xuesheng gaoji hanyu tingli xunlian zhong de wenti ji duice (浅谈越南学生高级汉语听力训练中的问题及对策) [Breve discusión sobre los problemas y las contramedidas en el entrenamiento avanzado de escucha del chino para estudiantes vietnamitas]”. *Xiandai Yuwen (现代语文) [Lenguas modernas de chino]*, (2), 88-89.
- Duan Junyu (段珺予). (2022). Problems and countermeasures while training the students of the international humanitarian university (odessa, ukraine) in listening comprehension at the intermediate level. *Actual problems of philology and professional training of specialists in multicultural space: an international journal*, (5), 130. Harbin: Harbin University of Engineering.
- Esnaola, L., y Ramón, H. (2023). Transformando la enseñanza con grandes modelos de lenguaje: una experiencia de utilización de la inteligencia artificial en el aula. *IV Workshop de innovación y transformación educativa*, Universidad Nacional del Noroeste de la Provincia de Buenos Aires. URL: <http://repositorio.unnoba.edu.ar/xmlui/handle/23601/651>
- Franciscu, S. (2023). “ChatGPT: A natural language generation model for Chatbots”. Working paper. DOI:10.13140/RG.2.2.24777.83044

- Gatica-Lara, F., y Uribarren-Berrueta, T. D. N. J. (2013). ¿Cómo elaborar una rúbrica?. *Investigación en educación médica*, 2(5), 61-65.
- Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu Jia, y Mirjalili, S. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. *Authorea Preprints*. 5-6.
- He Wanying (贺婉莹) y Han Lu (韩鹭). (2024). “Hanyu Shuiping Kaoshi tingli dui zhongwen xuexi de fanbo xiaoying yanjiu (HSK 听力考试对中文学习的反拨效应研究) [Investigación sobre el efecto de retorno de la prueba de escucha HSK en el aprendizaje del chino]”. *Hanyu Wenhua (汉语文化) [Cultura del chino]*, 350, 174-178. DOI:10.14014/j.cnki.cn11-2597/g2.2024.01.007.
- Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1), 38-53.
- Huang Shanshan. (2023). *Aplicación del ChatGPT en el aprendizaje de español como lengua extranjera*. (Trabajo de final de máster), Universidad de Barcelona.
- Jiaoyubu Zhongwai Yuyan Jiaoliu Hezuo Zhongxin (教育部中外语言交流合作中心). Zhongwen Shuiping Kaoshi HSK Si Ji Yangjuan (中文水平考试 HSK 四级样卷) [Documento de muestra para la prueba de dominio del chino HSK nivel 4]. Recuperado de: [https://download.chinesetest.cn/newhsk-site/model/HSK4\\_yj.pdf](https://download.chinesetest.cn/newhsk-site/model/HSK4_yj.pdf) [Fecha de consulta: 27-5-2024].
- Jin Lixin (金立鑫). (1998). “Tingli ke jiaoxue de xianzhuang yu gexin (听力课教学的现状与革新) [Situación actual e innovación de la enseñanza de las clases de escucha]”. *Duiwai Hanyu Luncong (对外汉语论丛) [Serie sobre la enseñanza del chino como lengua extranjera]*, 117-131. Shanghai Waiyu Jiaoyu Chubanshe

(上海外语教育出版社) [Prensa de educación de lenguas extranjeras de Shanghai].

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M., Weller, J., Kuhn, J., y Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, 103, 102274.

Ko, C. C. C. (2016). *Modelos experimentales de enseñanza y aprendizaje del chino como lengua extranjera a través de tareas y planteamientos competenciales* (Tesis doctoral), Universidad Pompeu Fabra.

Kohnke, L., Moorhouse, B. L., y Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal*, 00336882231162868. 3-9.

Lan, Yu-Ju, y Liao, Chia-Ying. (2018). The effects of 3D immersion on CSL students' listening comprehension. *Innovation in Language Learning and Teaching*, 12(1), 35-46.

Li Xiaying, Li, Belle, y Cho, Su-Je (2023). Empowering Chinese language learners from low-income families to improve their Chinese writing with ChatGPT's assistance afterschool. *Languages*, 8(4), 238.

Liang Wenjiao (梁文娇). (2015). "Xin Hanyu Shuiping Kaoshi si ji fudao jiaocai tingli bufen yanjiu (新 HSK 四级辅导教材听力部分研究) [Investigación sobre la parte auditiva del nuevo libro de texto tutorial HSK nivel 4]" (Trabajo de final de máster), Universidad Normal de Hebei (China).

Liddy, E. D. (2001). Natural language processing. *Encyclopedia of Library and Information Science*, 2nd Ed. NY. Marcel Decker, Inc. 3.

- Liu Songhao (刘颂浩). (2001). “Duiwai hanyu tingli jiaoxue yanjiu shuping (对外汉语听力教学研究述评) [Una revisión de la investigación sobre la enseñanza de la comprensión oral del chino como lengua extranjera]”. *Shijie Hanyu Jiaoxue (世界汉语教学) [Enseñanza del chino mundial]*, 1, 93-107.
- Ma Xiuli, Gong Yang, Gao Xuesong, y Xiang Yiqing. (2017). The teaching of Chinese as a second or foreign language: A systematic review of the literature 2005–2015. *Journal of Multilingual and Multicultural Development*, 38(9), 815-830.
- Meyer, J. G., Urbanowicz, R. J., Martin, P. C., O’Connor, K., Li, R., Peng, P. C., Bright, T. J., Tatonetti, N., Won, K. J., Hernandez, G. G., y Moore, J. H. (2023). ChatGPT and large language models in academia: opportunities and challenges. *BioData Mining*, 16(1), 20.
- Morales-Chan, M. A. (2023). Explorando el potencial de Chat GPT: Una clasificación de Prompts efectivos para la enseñanza. Working paper. <http://biblioteca.galileo.edu/tesario/handle/123456789/1348>
- Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., y Mian, A. (2023). A comprehensive overview of large language models. *arXiv preprint arXiv:2307.06435*. 19-29.
- Nazir, A., y Wang, Z. (2023). A comprehensive survey of ChatGPT: Advancements, applications, prospects, and challenges. *Meta-radiology*, 100022. 1-7.
- Paoliello, Antonio (2023). Evaluación y certificación, semana 2.1. ¿Cómo elaborar pruebas de idiomas? “Tingli lijie (听力理解) [Comprensión oral]” Comprensión auditiva/oral/audiovisual (CO). Documento inédito de clase.

- Paris, C. L., Swartout, W. R., y Mann, W. C. (Eds.). (2013). Natural language generation in artificial intelligence and computational linguistics. 119, 19-88. Springer Science & Business Media.
- Ram, B., y Verma, P. (2023). Artificial intelligence AI-based Chatbot study of ChatGPT, Google AI Bard and Baidu AI. *World Journal of Advanced Engineering Technology and Sciences*, 8(01), 258-261.
- Salguero Romero, P. (2023). La traducció pedagògica i l'ús de ChatGPT-3 a classes d'anglès com a segona llengua per a nens i nenes. (Trabajo de fin de grado), Universidad Autónoma de Barcelona.
- Shen Yan (沈燕). (1998). “Tan hanyu tingli jiaoxue zhong de zhendui xing (谈汉语听力教学中的针对性) [Sobre la pertinencia de la enseñanza de la escucha en chino]”. *Duiwai Hanyu Jiaoxue Tantaot Ji (对外汉语教学探讨集) [Colección de debates sobre la enseñanza del chino como lengua extranjera]*, 142-154. Beijing Daxue Chubanshe (北京大学出版社) [Prensa de la Universidad de Pekín].
- Shi Weidong (史卫东). (2000). “Kedong shuru yuanze yu zhongji hanyu tingli jiaocai shengci de chuli (可懂输入原则与中级汉语听力教材生词的处理) [El principio de entrada comprensible y procesamiento de nuevas palabras en libros de texto de comprensión auditiva en chino intermedio]”. *Huawen Jiaoxue yu Yanjiu (华文教学与研究) [Enseñanza e investigación del chino]*, (3).
- Singh, Ashish. (2022). Google Bard vs ChatGPT: How Google is taking on OpenAI with its new software, key differences. Recuperado de: <https://english.jagran.com/technology/bard-vs-chatgpt-how-google-is-taking-on-openai-with-its-new-software-key-differences-10064291> [Fecha de consulta: 24-2-2024].

- Singh, S. K., Kumar, S., y Mehra, P. S. (2023). Chat GPT & Google Bard AI: A Review. In 2023 International Conference on IoT, Communication and Automation Technology (ICICAT), 1-6. IEEE.
- Tai, T. Y., y Chen, H. H. J. (2021). The impact of immersive virtual reality on EFL learners' listening comprehension. *Journal of Educational Computing Research*, 59(7), 1272-1293.
- Vicente, M., Barros, C., Peregrino, F. S., Agulló, F., y Lloret, E. (2015). La generación de lenguaje natural: análisis del estado actual. *Computación y Sistemas*, 19(4), 721-756.
- Villalba, C. P. (2009). La evaluación de la comprensión oral en el aula de ELE. *MarcoELE. Revista de didáctica español lengua extranjera*, (9), 1-27.
- Wang Chunyu (2022). The Comparison between the Use of Audio-Visual Methods and Audio-only Methods in Improving Learners' Listening Proficiency in Chinese as a Second Language Classrooms. *Journal of Liberal Arts Prince of Songkla University*, 14(2), 181-205.
- Yan Lixiang, Sha Lele, Zhao Linxuan, Li Yuheng, Martinez-Maldonado, R., Chen Guanliang, Li Xinyu, Jin Yueqiao, y Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1), 90-112.
- Yu Na (于娜). (2007). “Duiwai hanyu tingli jiaoxue zongshu (对外汉语听力教学综述) [Una revisión de la enseñanza de la escucha del chino como lengua extranjera]”. *Liaoning Xingzheng Xueyuan Xuebao (辽宁行政学院学报) [Revista del Instituto de Administración de Liaoning]*, (12), 12.
- Zanón, N. T. (2010). Claves para comprender la destreza de la comprensión oral en lengua extranjera. *Epos*, (26), 195-216.

Zhang Jinjun (张晋军), Xie Nini (解妮妮), Wang Shihua (王世华), Li Yanan (李亚男) y Zhang Tieying (张铁英). (2010). “Xin Hanyu Shuiping Kaoshi yanzhi baogao (新汉语水平考试研制报告) [Nuevo informe de desarrollo de la prueba de dominio del chino]”. *Zhongguo Kaoshi (中国考试) [Examen de China]*, 9.

Zhongwen Kaoshi Fuwu Wang (中文考试服务网) [Red de servicios de exámenes de chino]. HSK Zhongwen Shuiping Kaoshi (HSK 中文水平考试) [La prueba de dominio del chino HSK]. Recuperado de: <https://www.chinesetest.cn/HSK> [Fecha de consulta: 5-2024].

Zhou Jie, Ke Pei, Qiu Xipeng, Huang Minlie, y Zhang Junping (2023). ChatGPT: potential, prospects, and limitations. *Frontiers of Information Technology & Electronic Engineering*, 1-6.