



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Application of Artificial Intelligence in

German Language Acquisition

- Evaluating ChatGPT's Ability to Generate Texts
as a Source of Comprehensible Input -

Luca Pellegrini

Department of German Language and Literature

The Graduate School of Sungshin University

Application of Artificial Intelligence in German Language Acquisition

- Evaluating ChatGPT's Ability to Generate Texts
as a Source of Comprehensible Input -

A Master's Thesis

Submitted to the

Graduate School of Sungshin University

in partial fulfillment of the requirements

for the degree of

Master of Arts

Luca Pellegrini

June, 2024

This is to certify that we have examined the
Master's Thesis of
Luca Pellegrini
Submitted to Department of German Language and Literature

Approved as to style and content:

Thesis Advisor: Hwang, Soonha (signature)
Committee Chairman: Kim, Gil-Ung (signature)
Committee Member: Kim, Gil-Ung (signature)
Committee Member: Kim, You-Ri (signature)
Committee Member: Hwang, Soonha (signature)

The Graduate School of Sungshin University

ABSTRACT

Application of Artificial Intelligence in German Language Acquisition - Evaluating ChatGPT's Ability to Generate Texts as a Source of Comprehensible Input -

Luca Pellegrini

Department of German Language and Literature

The Graduate School of Sungshin University

This research investigates the feasibility of using ChatGPT to generate texts at the A2 proficiency level to support German language acquisition through comprehensible input. The thesis is structured into five sections. The second section examines the theoretical framework, focusing on Input theory and the application of AI in education. It analyzes the distinction between language acquisition and learning, delves into the Input Hypothesis, and emphasizes the role of reading in receiving comprehensible input. The section also explores the technological aspects of ChatGPT, its functionality, and its potential in creating educational content, while addressing the challenges and limitations of AI in educational contexts. The third section introduces the empirical component, outlining criteria to assess the effectiveness of texts generated by ChatGPT, including text length adaptability, readability, and vocabulary appropriateness according to the Common European Framework of Reference for Languages. The fourth section applies these criteria to evaluate whether ChatGPT can consistently produce texts that meet the A2

proficiency level in German. The final section synthesizes the results from the experimental analysis, evaluating the practical implications of using AI-generated texts in language learning. The conclusion provides insights into the applicability of AI tools like ChatGPT in enhancing the language acquisition process, informed by both theoretical and empirical evidence.

Keywords: ChatGPT, German language acquisition, comprehensible input, Input Hypothesis, text generation

Table of contents

1. Introduction	1
2. Theory	4
2.1. Language Acquisition within the Framework of Krashen's Theory	4
2.1.1 Foreign Language Acquisition and Learning	5
2.1.2. The Input Hypothesis and the Role of Comprehensible Input	6
2.1.3. The Power of Reading in Foreign Language Acquisition	9
2.1.4. Conclusion	12
2.2. ChatGPT and Education	13
2.2.1. Overview of ChatGPT	14
2.2.2. Application of ChatGPT in Education	16
2.2.3. Risks and Limitations of Using ChatGPT in Education	18
2.2.4. Conclusion	20
3. Methodology	22
3.1. Length	27
3.2. Readability	32
3.3. Vocabulary	37
3.4. Conclusion	41
4. Experiment and Evaluation	43
4.1. Texts without controlled length	49
4.1.1. Length	49
4.1.2. Readability	50
4.1.3. Vocabulary	52

4.2. Texts with controlled length	54
4.2.1. Length	54
4.2.2. Readability	56
4.2.3. Vocabulary	58
4.3. Conclusion	60
5. Conclusion	62
Bibliography	65

List of Tables

Table 1: Common European Framework of Reference for Languages (CEFR)	22
Table 2: Texts (Goethe-Zertifikat)	25
Table 3: Length (Goethe-Zertifikat A2)	30
Table 4: Readability (Goethe-Zertifikat A2)	36
Table 5: Vocabulary (Goethe-Zertifikat A2)	39
Table 6: Prompts	43
Table 7: Texts (ChatGPT)	44
Table 8: Length (ChatGPT without controlled length)	49
Table 9: Readability (ChatGPT without controlled length)	51
Table 10: Vocabulary (ChatGPT without controlled length)	53
Table 11: Length (ChatGPT with controlled length)	55
Table 12: Readability (ChatGPT with controlled length)	56
Table 13: Vocabulary (ChatGPT with controlled length)	58

List of texts

Text 7 (Goethe-Zertifikat)	27
Text 8 (Goethe-Zertifikat)	29
Text 9 (Goethe-Zertifikat)	33
Text 10 (Goethe-Zertifikat)	34
Text 11 (Goethe-Zertifikat)	38
Text 4 (ChatGPT)	46
Text 2 (ChatGPT)	47

1. Introduction

Learning German, with its intricate syntax and rich vocabulary, presents a significant challenge for many students. Fortunately, this challenge might be alleviated by the availability of high-quality educational material, such as those found within textbooks or across various online platforms. However, most high-quality resources are not free, and the free ones, such as newspaper articles or blogs, often prove too challenging for beginners. Additionally, much of the available learning material covers a limited range of topics, which may not capture the interest of all students aiming to learn German. This situation raises an important question: could there be a way to generate customizable texts at no cost, allowing learners to choose their topics of interest?

To address this issue, it may be feasible to harness artificial intelligence, enabling students to receive comprehensible input tailored to their learning needs. The following research investigates the application of artificial intelligence in the context of foreign language acquisition, with a particular focus on the capabilities of ChatGPT, the advanced language model developed by OpenAI. The starting point of this study is Stephen Krashen's Input Hypothesis from his broader theory of second language acquisition, which asserts that effective language learning occurs through exposure to language input that is just slightly beyond the learner's current proficiency level. According to Krashen, such comprehensible input is critical for advancing from one stage of knowledge to the next, facilitating the natural acquisition process (Krashen, 1989). This research aims to determine whether it is possible to use ChatGPT to generate texts that serve as

comprehensible input, effectively supporting the acquisition of the German language.

The relevance of this research lies in its exploration of modern technological tools that can support traditional language learning methodologies. This study aims to explore the extent to which ChatGPT can contribute to foreign language acquisition by assessing its ability to generate German language texts that serve as comprehensible input. Indeed, previous studies have primarily explored ChatGPT's impact on improving specific language skills such as enhancing students' writing proficiency (Hanane, 2023), improving lexico-grammatical features (Abdelhadi, 2023), academic writing (Dawa, Dhendup, Tashi, Rosso 2023), and paraphrasing texts (Amini-Salehi, Hassanipour, Bozorgi et al, 2023). This research, however, investigates ChatGPT's potential as a source of comprehensible input for German learners.

The structure of this thesis is divided into five sections. The second section deals with the theoretical framework of this research, namely Krashen's theory and the application of AI for educational purposes. First of all, this section provides an analysis of Krashen's theory, specifically distinguishing between the concepts of language acquisition and learning. It further explores the Input Hypothesis in detail, elucidating how language acquisition is facilitated through the mechanism of comprehensible input. This section also highlights the importance of reading as a potent method for receiving comprehensible input, which is essential for effective language acquisition. Moreover, technological aspects of this study will be taken into account, detailing the functionalities and mechanics of ChatGPT. It provides an overview of artificial intelligence in educational contexts, outlining how AI can

be optimized to generate educational content. This includes a discussion of the potential educational applications of ChatGPT, focusing on its capacity to generate language learning materials that are engaging and pedagogically valuable. The limitations and challenges of employing AI in educational settings, particularly regarding the authenticity and contextual appropriateness of the content generated, will also be critically examined.

The third section presents the empirical component of the research, by introducing the criteria to assess the effectiveness of texts generated by ChatGPT. The evaluation criteria include the adaptability of text length, the readability of the texts, and the appropriateness of vocabulary as per the Common European Framework of Reference for Languages.

In the fourth section these criteria will be used in order to determine whether the AI-generated texts meet the ChatGPT can consistently generate texts that meet the A2 proficiency level in German in terms of word length, readability, and vocabulary. This analysis will help establish ChatGPT's capability in producing educational material that aligns with specific learning benchmarks for the A2 proficiency level in German.

The final part of the thesis synthesizes the results from the experimental analysis and evaluates the practical implications of using AI-generated texts in language learning environments. The conclusion aims to offer comprehensive insights into the applicability of AI tools like ChatGPT in enhancing the language acquisition process, considering both the theoretical framework and empirical evidence gathered throughout the study.

2. Theory

The quest to uncover how language acquisition works has long been a subject of scholarly inquiry and practical interest. The first part of this section delves into language acquisition within the framework of Krashen's Theory. It begins with an examination of the fundamental distinction between acquisition and learning. Subsequent analysis focuses on the Input Hypothesis, which claims that acquiring a foreign language is achievable through exposure to comprehensible input. The discussion also covers the significant role of reading in foreign language acquisition. The second part of this section investigates the application of ChatGPT within educational contexts. It assesses the potential benefits and applications of ChatGPT in education, alongside potential negative impacts that need careful consideration. This analysis aims to elucidate the broader implications of incorporating advanced AI technologies like ChatGPT into educational frameworks.

2.1. Language Acquisition within the Framework of Krashen's Theory

The topic of foreign language acquisition has been extensively explored by many scholars, yet this research will specifically focus on the contributions of Stephen Krashen. His theory highlights the crucial role of comprehensible input in foreign language acquisition and emphasizes the importance of reading as a means to obtain this input. By examining Krashen's principles, this study aims to assess whether texts generated by ChatGPT can effectively contribute to

enhancing language skills. This approach will provide a focused analysis on the practical applications of comprehensible input in language learning and the potential of AI-generated texts to facilitate this process. Krashen's theory suggests that language acquisition occurs most effectively when learners are exposed to comprehensible input, a principle that underpins the foundation of this thesis. Before dealing with the use of ChatGPT in language acquisition, it is important to analyze key components of Krashen's theory. This section starts with an exploration of the fundamental concepts of acquisition and learning, delineating the crucial distinction of Krashen's theoretical framework. Second, it is possible to delve into the Input Hypothesis, elucidating the mechanisms through which language acquisition occurs and the role of comprehensible input. Furthermore, I shall elucidate the criteria for optimal input, essential for discerning the types of linguistic stimuli conducive to language acquisition. Finally, this study examines the significance of reading as a means of encountering comprehensible input in the process of acquiring a second language. First of all, it is necessary to have an overview of ChatGPT and, in more general terms, of Artificial Intelligence.

2.1.1 Foreign Language Acquisition and Learning

The cornerstone of Krashen's theory lies in the differentiation between acquisition and learning. Krashen claims that adults employ two distinct and autonomous methods in developing their language abilities: acquisition and learning. Acquisition reflects the natural process observed in children as they develop proficiency in their first language. This process operates primarily at a

subconscious level, with the acquirer often unaware of the ongoing acquisition process and the derived linguistic competence stored within the brain. Such a process leads to the formation of a tacit understanding of language.

Conversely, learning refers to the conscious acquisition of language knowledge, including grammar rules and linguistic rules. In the context of second language acquisition, it appears that acquisition is more effective compared to learning, although learning does serve a vital function as an editor, facilitating corrections based on the grammatical principles learned. Nonetheless, Krashen indicates that fluency and accuracy in second language performance predominantly stem from acquired rather than learned language skills. While error correction may aid learners in deducing language rules, its impact on the subconscious acquisition of language competence remains minimal (Krashen, 1989:8).

2.1.2. The Input Hypothesis and the Role of Comprehensible Input

Given the fundamental role of acquisition in language competence development, it is imperative to delve into how language is concretely acquired. Krashen proposes the Input Hypothesis, claiming that acquiring a foreign language is possible through exposure to comprehensible input (Krashen, 1989:9). In other words, in order to move from one learning stage (i) to a stage that the acquirer has not yet acquired but is developmentally ready to acquire ($i + 1$), the acquirer must understand input containing elements from $i + 1$.

In terms of vocabulary acquisition, learners can acquire new words when they understand what they read or hear by making predictions about the meaning of

input when it contains new aspects of language. These predictions often rely on world knowledge, language understanding, and familiarity with the writing system. Consequently, as learners encounter and grasp a specific word in different contexts, they gradually understand its full meaning and grammatical usage. Hence, texts including a high percentage of known language and dealing with content familiar to the reader are optimal for language development, as they facilitate accurate predictions. In this regard, Laufer, for example, suggests that vocabulary development takes place when 95 percent or more of the text's vocabulary is already acquired (Laufer, 1992).

This perspective contrasts sharply with theories like the Skill-Building Hypothesis, which advocates for mastering grammar and vocabulary before language acquisition. Despite the apparent paradox of understanding yet-to-be-acquired elements, it is crucial to acknowledge the role of context and extralinguistic elements in facilitating comprehension. Successful communication occurs when input is comprehensible and provided effortlessly.

Moreover, according to the Input Hypothesis, speaking ability emerges indirectly through exposure to comprehensible input as a byproduct of language acquisition. This aligns with observations in first language acquisition, where children often undergo a silent period before vocalizing. Thus, with sufficient high-quality comprehensible input, it is possible to be exposed to all the structures and eventually acquire them.

However, it is essential to note that comprehensible input alone is not adequate for second language acquisition. Dulay and Burt introduce the concept of an Affective Filter, which hampers effective input processing by the Language Acquisition

Device. Factors such as low motivation, high anxiety, and diminished self-esteem contribute to the Affective Filter (Dulay & Burt, 1977). Consequently, acquisition thrives when learners are in low-anxiety environments exposed to compelling and comprehensible input.

Despite the presence of the Affective Filter, the research underlines the fundamental role of comprehensible input in foreign language acquisition (Krashen, 1982). Krashen claims that optimal input should be comprehensible, compelling, ungrammatically sequenced, and provided in sufficient quantity.

Foremost among these characteristics is comprehensibility. It is virtually impossible to acquire a foreign language solely through passive activities like watching television or listening to the radio if the content is beyond the acquirer's current understanding. Particularly at beginner levels, comprehension can be enhanced by exposure to speech delivered at a slower pace with clear articulation, aiding in the identification of word boundaries. Additionally, the use of high-frequency vocabulary and simplified syntax can enhance understanding, as often observed in caretaker speech, foreigner talk, and teacher talk (Hatch, 1979). Secondly, optimal input should be engaging to ensure the acquirer's attention. Ideally, the content should be so captivating that the acquirer forgets they are exposed to content in another language. Moreover, such input should be similar to the content they would eagerly consume in their native language. This explains why mechanical drills aimed solely at practicing grammatical structures may be ineffective in language acquisition (Paulston, 1972). However, it might be challenging to find material that balances comprehensibility with interest. Indeed, a great amount of comprehensible input in textbooks may lack appeal, while

compelling content may be beyond the acquirer's level of proficiency. Nevertheless, finding such engaging input is crucial, as evidenced by cases where individuals made significant progress in a foreign language simply by immersing themselves in captivating books or films, without explicit intent to improve their language skills.

Thirdly, input should not be grammatically sequenced. This means that acquirers should not deliberately include the stage beyond their current level of competence ($i + 1$) in the input they are exposed to because when input is comprehensible, $i + 1$ will naturally occur. This contradicts the approach often found in textbooks, which meticulously follow the sequence of precise grammatical rules. However, the concern with such textbooks lies in their focus on grammar-based programs, which can reduce the quality of comprehensible input and overlook communicative objectives.

Lastly, optimal input must be provided in sufficient quantity. While determining the exact quantity necessary for proficiency remains challenging, it is essential to ensure learners receive a considerable amount of input. In this case, it is important to avoid artificial constraints on input quantity, with special emphasis placed on content-rich exposure to expedite and enhance language acquisition.

2.1.3. The Power of Reading in Foreign Language Acquisition

In this regard, Krashen claims that reading serves as an excellent source of comprehensible input, particularly when aligned with the aforementioned key principles comprehensibility and interest (Krashen, 2011: 70). Engaging with texts

in a foreign language facilitates vocabulary expansion, enhances writing proficiency, and refines grammatical competence. Evidence indicates that reading exposure alone significantly influences language development, especially in areas such as reading comprehension, writing style, and the acquisition of sophisticated vocabulary and grammar (Krashen, 2004). For instance, the study by Stokes, Krashen, and Kartchner (1998) highlights the positive correlation between the amount of voluntary reading and the accurate usage of the subjunctive mood in Spanish. Similarly, it has been possible to establish a connection between extensive reading and performance on assessments like the TOEFL examination, the test of academic English in reading and listening comprehension, grammar, and writing (Constantino, Lee, Cho & Krashen, 1997). This trend extends to first language acquisition, where individuals reporting higher reading frequency typically demonstrate superior reading skills and exhibit a more mature writing style.

Moreover, narrow reading also proves effective in language acquisition. Narrow reading is the type of reading focusing on one topic, author, or genre, according to the reader's interests, in order to expand the amount of reading over time. Such type of reading not only sustains the reader's interest but also enhances comprehensibility due to increased background knowledge. Krashen holds that individuals who engage in narrow reading make significant progress in language proficiency, as deeper engagement with content leads to exposure to diverse syntactic structures and specialized vocabulary beyond the chosen topic (Krashen 2004).

According to Krashen (2017), in order to achieve advanced proficiency in a foreign

language, three stages are essential, each emphasizing compelling comprehensible input. The initial stage involves exposure to stories or read-alouds, fostering linguistic competence and sparking interest in books. Subsequently, free voluntary reading contributes to building linguistic and knowledge foundations necessary for academic reading comprehension. In the final stage, engaging with specialized materials improves academic linguistic competence.

In this regard, Krashen formulated various hypotheses that shed light on the importance of being exposed to input by reading in a foreign language (Krashen, 2011: 81):

a) The Forgetting Hypothesis suggests that language acquisition occurs when readers are so interested in the content that they forget it is in another language, promoting automatic skill development. This aligns with the Pleasure Hypothesis, which claims that language acquisition is most effective when learners find the activity enjoyable, leading to a state of “flow” where learners are deeply engaged but effortlessly involved.

b) The Effortless Reading Hypothesis underlines the importance of high comprehension levels in optimal reading input for language development. This theory aligns with the aforementioned assertion that, in order to develop vocabulary, at least 95 percent of the words in a text must be familiar (Laufer 1992).

c) Similarly, the Unawareness of Acquisition Hypothesis suggests that readers may not realize their progress during engaging reading experiences, as evidenced by cases where learners express surprise at their acquired knowledge.

d) The Comprehension Checking Hypothesis argues that the more acquirers check comprehension, the less readers understand and the less they acquire, because checking implies less engagement, less flow, and less enjoyment of reading. Hence, language development occurs when learners focus on content rather than language itself.

2.1.4. Conclusion

In conclusion, this section has delved into the fundamental aspects of Krashen's theory. Firstly, it has been elucidated that the central distinction lies between acquisition and learning. Acquisition, being a subconscious process, progresses without the acquirers being consciously aware of their development, while learning refers to conscious knowledge of the language, including grammar and its rules. While learning may serve as an editor, facilitating corrections during language acquisition, it has been demonstrated that true fluency stems from acquisition, even among individuals with a highly analytical approach to learning. Furthermore, a detailed analysis of the Input Hypothesis has been conducted, which posits that language acquisition occurs through exposure to comprehensible input. Thus, for the acquisition of new vocabulary, it is imperative to understand the meaning of what is read or heard, often through making predictions when encountering unfamiliar language elements. The development of speaking skills, rather than being directly taught, emerges organically through exposure to comprehensible input as a consequence of language acquisition. Effective language acquisition, therefore, necessitates exposure to input that is not only

comprehensible but also engaging, devoid of strict grammatical sequencing, and provided in sufficient quantity. Hence, it has been established that reading serves as an exceptional source of comprehensible input. Indeed, language acquisition is most effective when individuals are deeply interested in the content they are reading to the extent that they forget it is in a foreign language, and when the material is easy to comprehend. Through this immersive process, readers can acquire a foreign language without consciously realizing the progress they are making. The conclusion drawn from Krashen's theory is of great importance as it directs the trajectory of my future research endeavors. After having elucidated how language acquisition works, which features the linguistic input should have, and which role reading plays in language development, it will be possible to analyze the generation of texts by ChatGPT.

2.2. ChatGPT and Education

After having analyzed the foundation laid in Krashen's theory, it is possible to delve into ChatGPT and its applicability in educational contexts. First of all, it is imperative to give an overview of ChatGPT and artificial intelligence in general, thus elucidating the mechanics of ChatGPT, delineating the spectrum of materials it can generate, and introducing optimal strategies for formulating prompts to use the language model effectively. Moreover, it is important to examine the different applications of ChatGPT, with a specific focus on its educational uses. In this regard, I shall explain how ChatGPT could potentially facilitate the language acquisition process, particularly in the realm of learning foreign languages, in order

to understand the transformative potential of ChatGPT in educational settings. However, it is equally crucial to acknowledge the inherent limitations associated with employing ChatGPT for educational purposes. In this way, it will be possible to delineate the true potential of ChatGPT in facilitating language acquisition processes, while also delineating its boundaries and challenges within educational contexts.

2.2.1. Overview of ChatGPT

First of all, it is necessary to have an overview of ChatGPT and, in more general terms, of Artificial Intelligence. The concept itself of “Artificial Intelligence” (AI) includes various techniques and methodologies aimed at creating intelligent machines capable of perceiving their environment and taking appropriate actions. The rapid development of AI has led some to regard 2023 as the “Year of Artificial Intelligence”, given its profound impact on societal life (Mercedes, Olite, Del et al, 2023: 3). Nowadays, AI, in particular Generative AI, is revolutionizing how businesses and individuals interact with technology. Generative AI involves systems that can generate new data or outputs, such as images, music, or text, by processing and generating natural language text. Natural Language Processing (NLP) is also a fundamental component of AI, enabling machines to understand and respond to human speech (Mhlanga, 2023). This is the case for GPT (Generative Pre-trained Transformer), a type of language model developed by OpenAI that uses deep learning algorithms to generate human-like responses to text-based prompts (Božić & Poola, 2023).

GPT models, such as GPT-3.5 and GPT-4, have been trained on vast amounts of text data from the internet. While GPT-3.5 is text-based, accepting and producing text output, GPT-4 incorporates elements of computer vision, enabling it to process images in addition to text (Marquez, 2024). These models have been subjected to various tests, including bar exams. It has been established that GPT-3.5's score was around the bottom 10% of the test takers in the exam whereas GPT-4's score is more around the top 10% of test takers (Gimpel, Hall & Decker, 2023: 15). The essence of GPT lies in its ability to understand, generate coherent responses to natural language inputs, and incorporate self-attention mechanisms to consider contextual information, ensuring a conversational flow. Moreover, these models improve their performance over time by learning continuously from user interactions.

In order to optimize interaction with GPT, well-formulated prompts prove to be essential (Morales-Chan, 2023). Prompts serve as directives for the language model to generate responses. Indeed, when interacting with ChatGPT, it is essential to provide clear and precise prompts, avoiding vague or confusing inquiries that may result in imprecise answers. In general, prompts must contain the topic, the style, the tone, the context, and any necessary contextual information. When formulating prompts, it is fundamental to define a clear objective and maintain conciseness. Furthermore, using natural language and providing clear definitions for technical terms enhance understanding, and a careful selection of verbs guides the AI's actions and influences the accuracy and relevance of responses. Choosing the appropriate verb ensures the AI comprehends the task at hand, leading to more precise solutions and

recommendations. Hence, effective interaction with a large language model (LLM) like GPT requires well-formulated prompts that provide clear context, concise objectives, and appropriate verb usage, thus fostering coherent and relevant responses.

2.2.2. Application of ChatGPT in Education

After having been introduced in November 2022, ChatGPT has acquired popularity very rapidly, boasting over one million users within its first five days (Mercedes, Olite, Del et al, 2023: 6). Over the following two years, it has established itself as an invaluable tool for people, companies and organizations. Indeed, ChatGPT excels in diverse tasks, from composing reports and email messages to summarizing, translating, scripting, and coding, and it has thus become indispensable across industries such as education, healthcare, customer support, and more (Mhlanga, 2023). In healthcare, for instance, ChatGPT is able to improve patient outcomes through the analysis of extensive medical data in diagnoses and treatments. Similarly, in finance, it streamlines processes, reduces costs, and enhances transaction speed and accuracy. Its potential to revolutionize these sectors lies in its ability to process vast amounts of data very quickly and automate routine tasks, thereby enhancing efficiency and productivity. Despite being in a developmental phase and susceptible to errors, ChatGPT's quality responses have pleasantly surprised users, which has contributed to its rising popularity. Its user-friendly interface simplifies the application, making it accessible even to those unfamiliar with AI systems.

Beyond understanding and responding to context, ChatGPT can also generate creative content like fiction, poetry, and music, presenting novel possibilities for education (Božić & Poola, 2023). Indeed, in higher education, ChatGPT can enrich learning experiences by facilitating interactive language practice, instant feedback, and personalized learning materials. Its ability to simulate real-life conversations contributes to language acquisition, while its feedback on grammar and writing enhances students' writing skills (Mhlanga, 2023). By generating prompts and questions for discussions, ChatGPT develops students' critical thinking and enables the exploration of diverse perspectives (Benavides, 2023). Moreover, ChatGPT takes into account individual needs and proficiency levels, adapting content complexity in order to optimize learning outcomes. Its personalized vocabulary development strategies and individualized learning materials further enhance engagement and motivation, ultimately leading to an improvement in language learning results. Thus, it can be claimed that ChatGPT's potential to transform education lies in its capacity to provide tailored, interactive learning experiences that allow students to develop language proficiency and critical thinking skills effectively. With its ability to adapt to individual needs and deliver instant feedback, ChatGPT paves the way for a more engaging and impactful learning journey.

In this context, it must be pointed out that the existing research on ChatGPT has primarily focused on enhancing students' writing proficiency (Hanane, 2023), improving lexico-grammatical features (Abdelhadi, 2023), academic writing (Dawa, Dhendup, Tashi, Rosso 2023), and paraphrasing texts (Amini-Salehi, Hassanipour, Bozorgi et al, 2023). However, there appears to be a gap in exploring its potential

for creating learning materials focused on reading. Additionally, there seems to be limited investigation into ChatGPT's ability to control language levels in foreign language learning contexts. My research aims to bridge these gaps, investigating the extent to which language levels can be managed in text production with ChatGPT. Indeed, by allowing acquirers to delve into topics that genuinely captivate their curiosity, ChatGPT has the potential to nurture intrinsic motivation and foster enthusiasm for learning. This, in turn, can improve students' engagement and perseverance, leading to more meaningful and enduring learning outcomes. Moreover, the acquirers can assume control over the selection of domains, topics, and questions, thus reaching a level of autonomy that diverges from the traditional language textbooks. If this capability can be effectively demonstrated, it could serve as a compelling impetus for further inquiry into the role of LLM models such as ChatGPT in text production research.

2.2.3. Risks and Limitations of Using ChatGPT in Education

While ChatGPT has the potential to bring many benefits to education, it also presents potential negative impacts that must be taken into account. Foremost among these concerns are social issues, particularly the impact on employment and privacy. The automation of jobs through AI technology could result in significant job losses and increased unemployment rates. Moreover, AI algorithms collect vast amounts of personal data, raising concerns about privacy and data security, including vulnerability to hacking and cyber-attacks.

In the educational context, there are additional risks related to the dependency on ChatGPT and other AI tools. Overreliance on these tools may decrease students' independence in learning, thus limiting originality in generating text (Božić & Poola, 2023). Furthermore, although ChatGPT can simulate conversations, it might not be able to replace the social and emotional interactions inherent in face-to-face communication, and the invaluable emotional support provided by teachers and their adaptability to individual student needs.

Another significant concern is the potential for inaccuracies and biases in ChatGPT responses, particularly in models like GPT 3.5 (Božić & Poola, 2023). Biases and inaccuracies in training data can propagate into generated text, thus leading to unintentional plagiarism, copyright infringements, or inaccuracies in quotations. Additionally, ChatGPT's proficiency is primarily in English, which may limit its effectiveness in other languages (Morales-Chan, 2023).

To mitigate these challenges, it is crucial to use ChatGPT as a supportive tool in education rather than a replacement for human teachers or critical thinking skills (Benavides, 2023). Promoting ethics and academic integrity among students is also essential in order to enable informed decision-making regarding the use of ChatGPT and other technological resources. One strategy that can be employed to address these limitations is to confine tasks involving creativity and critical reflection to face-to-face classroom settings. Here, students can engage in activities such as generating and discussing research ideas, writing essays, elaborating reflections, and developing projects and presentations under the guidance of educators.

However, it could be argued that the risks and limitations mentioned primarily refer to social and pedagogical considerations. They may not necessarily directly impact the generation of texts aimed at providing acquirers with comprehensible input, which stands as the primary focus of my research. Therefore, it might be prudent to consider setting aside these other aspects and maintaining focus solely on the core objective. Additionally, it could be suggested that the success of integrating ChatGPT technology into education might depend on thoughtful integration into the learning process to potentially enhance student learning and achievement. By adopting a balanced approach, educators might be able to develop the potential of ChatGPT to enrich educational experiences and empower students for the future.

2.2.4. Conclusion

In conclusion, I have dealt with ChatGPT and its potential applications in education, particularly in foreign language acquisition. As has been mentioned, ChatGPT, an advanced language model developed by OpenAI, uses deep learning algorithms to analyze vast text datasets, thus identifying linguistic patterns, and producing coherent, grammatically correct sentences. However, It is imperative to use correctly the prompts in order to harness the full potential of ChatGPT. This includes providing clear, precise context, maintaining specificity and conciseness, avoiding ambiguity, and employing exact verbs. Moreover, it has been observed that ChatGPT has already been used in diverse fields such as healthcare and business. Within the realm of education, particularly language learning, ChatGPT also exhibits immense potential, as this language model is able

to simulate authentic conversations, offer learners immediate feedback on grammar, pronunciation, and vocabulary usage, and even enhance their communicative competence and writing proficiency. Finally, I have addressed the inherent risks and limitations associated with ChatGPT, which are mainly related to ethical considerations and pedagogical concerns rather than limitations in generating learning materials that allow acquirers to be exposed to comprehensible input. Hence, it may be beneficial to consider conducting research to investigate whether ChatGPT has the potential to offer acquirers the comprehensible input proposed in Krashen's theory. This research could help assess the possibility of controlling the language level of generated texts to provide acquirers with comprehensible input, thus facilitating language acquisition. If ChatGPT proves effective in controlling language levels, it could provide foreign language learners with more accessible input compared to traditional textbooks. However, if challenges arise in controlling language levels, it may limit the usefulness of texts generated by ChatGPT for language learning purposes, despite the advantage of allowing learners to create texts aligned with their personal interests.

3. Methodology

The methodology section delves into an examination of the possibility of generating texts by using ChatGPT in order to provide language acquirers with comprehensible input. Within the framework of this research, an investigation will be conducted focusing on texts generated in the German language, specifically targeting proficiency level A2 of the Common European Framework of Reference for Languages (CEFR). It is important to note that in German, as in other European languages, proficiency levels are categorized into six levels, as can be seen in the following table:

Table 1: Common European Framework of Reference for Languages (CEFR)

Proficient User	C2	Can understand with ease virtually everything heard or read. Can summarise information from different spoken and written sources, reconstructing arguments and accounts in a coherent presentation. Can express him/herself spontaneously, very fluently and precisely, differentiating finer shades of meaning even in more complex situations.
	C1	Can understand a wide range of demanding, longer texts, and recognise implicit meaning. Can express him/herself fluently and spontaneously without much obvious searching for expressions. Can use language flexibly and effectively for social, academic and professional purposes. Can produce clear, well-structured, detailed text on complex subjects, showing controlled use of organisational patterns, connectors and cohesive devices.

Independent User	B2	Can understand the main ideas of complex text on both concrete and abstract topics, including technical discussions in his/her field of specialisation. Can interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party. Can produce clear, detailed text on a wide range of subjects and explain a viewpoint on a topical issue giving the advantages and disadvantages of various options.
	B1	Can understand the main points of clear standard input on familiar matters regularly encountered in work, school, leisure, etc. Can deal with most situations likely to arise whilst travelling in an area where the language is spoken. Can produce simple connected text on topics which are familiar or of personal interest. Can describe experiences and events, dreams, hopes and ambitions and briefly give reasons and explanations for opinions and plans.
Basic User	A2	Can understand sentences and frequently used expressions related to areas of most immediate relevance (e.g. very basic personal and family information, shopping, local geography, employment). Can communicate in simple and routine tasks requiring a simple and direct exchange of information on familiar and routine matters. Can describe in simple terms aspects of his/her background, immediate environment and matters in areas of immediate need.
	A1	Can understand and use familiar everyday expressions and very basic phrases aimed at the satisfaction of needs of a concrete type. Can introduce him/herself and others and can ask and answer questions about personal details such as where he/she lives, people he/she knows and things he/she

		has. Can interact in a simple way provided the other person talks slowly and clearly and is prepared to help.
--	--	---

(Council of Europe, 2001:24)

The choice to focus on level A2 within the CEFR framework is deliberate, as it represents a proficiency level characterized by basic language skills. By focusing on this proficiency level, the study seeks to evaluate the ability of ChatGPT to produce texts that meet acquirers' needs at the initial stages of language acquisition, where comprehensible input is crucial for effective learning. It should be noted that this research will not consider whether the texts generated by ChatGPT address the specified topic in the prompt, as ChatGPT's efficiency in producing topic-specific texts has already been dealt with in the second part of this research. Additionally, this analysis will not assess the accuracy of the information provided by ChatGPT, as this does not seem to impact foreign language acquisition. However, these aspects could be considered in future research endeavors.

In this context, it is imperative to consider materials at the A2 level provided by the Goethe Institut. Specifically, the Goethe-Zertifikat A2 Modellsatz Erwachsene and Goethe-Zertifikat A2 Modellsatz Jugendliche (Goethe-Institut, 2013), and the four exams found in Mit Erfolg zum Goethe Zertifikat A2 (Loumiotis & Näfken, 2016) might be used to this aim. The Goethe-Zertifikat A2 proficiency exam encompasses four parts in the reading section: part 1 features a newspaper article, part 2 an information board, part 3 an email, and part 4 various

announcements. Given the focus of my research on length and vocabulary, it might be reasonable to consider only parts 1 and 3, which offer longer texts rich in vocabulary. Hence, a total of 12 texts (6 newspaper articles and 6 emails) from 6 exams Goethe-Zertifikat A2 will be used within the framework of my research, as illustrated in this table:

Table 2: Texts (Goethe-Zertifikat)

Goethe-Zertifikat A2 Modellsatz Erwachsene		
Text	Genre	Topic
Text 1	newspaper article	TV show
Text 2	e-mail	Life and study in Hamburg
Goethe-Zertifikat A2 Modellsatz Jugendliche		
Text	Genre	Topic
Text 3	newspaper article	Youth Newspaper
Text 4	e-mail	Invitation to one's home
Mit Erfolg zum Goethe Zertifikat A2		
Text	Genre	Topic
Text 5	newspaper article	Example of integration
Text 6	e-mail	Class reunion
Text 7	newspaper article	Effects of coffee
Text 8	e-mail	Internship abroad
Text 9	newspaper article	Wild birds
Text 10	e-mail	Problems with girlfriend
Text 11	newspaper article	Remote work

Text 12	e-mail	Accident
---------	--------	----------

In order to evaluate the efficacy of ChatGPT in this regard, five criteria will be taken into account.

Firstly, the length of the generated texts will be examined to ascertain the extent to which it can be controlled or determined. Understanding the variability in text length generated by ChatGPT is essential for developing learning materials to meet specific instructional needs and preferences.

Secondly, the readability of the texts generated will be considered in order to ascertain whether they are sufficiently clear and understandable to serve as comprehensible input for language learners. This criterion is quite important as it directly impacts the effectiveness of the generated texts in facilitating language acquisition.

Finally, the vocabulary employed in the generated texts will be evaluated to determine if it aligns with the proficiency levels outlined in the Common European Framework of Reference for Languages. Assessing the correspondence of the vocabulary to established language proficiency standards provides insights into the suitability of the generated texts for learners at different proficiency levels.

While this analysis does not include factors such as grammar, syntax, and subjective criteria which may be challenging to evaluate, it might be assumed that by analyzing the aforementioned three criteria, it will be possible to establish the capabilities and limitations of ChatGPT in generating texts conducive to language acquisition.

3.1. Length

With regard to the first criterion, the length of texts generated by ChatGPT will be taken into consideration. It is essential to determine the extent to which the length of these texts can be controlled or determined. Given that the primary focus of this research is on generating texts for reading, it is reasonable to establish the appropriate length for texts at the A2 level. However, there is no precise information regarding the ideal length of texts for A2-level readers. While the Goethe Zertifikat A2, the proficiency exam for the German language at this level, provides guidelines for the writing section (20-30 words for the first part and 30-40 words for the second), no specific guidance is provided for the reading section. To determine the appropriate length for A2-level texts, I have analyzed parts 1 and 3 in the Goethe-Zertifikat A2 Modellsatz Erwachsene, in the Goethe-Zertifikat A2 Modellsatz Jugendliche (Goethe-Institut, 2013), and in Mit Erfolg zum Goethe Zertifikat A2 (Loumiotis & Näfken, 2016), and have calculated the total word count for the reading passages by counting the lemmas in the texts, as illustrated in the following examples.

Text 7 (Goethe-Zertifikat)

Ist Kaffee schädlich oder gesund?

„C-A-F-F-E-E, trink nicht so viel Caffee“...

... so lautet ein altes Lied, das vor den Gefahren des Kaffees warnen sollte.

Ist aber Kaffee tatsächlich so gefährlich? In Zeitschriften liest man oft

negative Artikel über Kaffee: Kaffee macht müde", „Kaffee macht krank", aber auch positive: „Kaffee macht munter", „Kaffee macht schön". Was davon richtig ist, darüber streiten sich die Experten.

Man weiß aber sicher, welchen Weg der Kaffee durch unseren Körper nimmt: Er geht schnell ins Blut und in unser Gehirn. Was macht aber dann das Koffein, das ja im Kaffee ist, mit uns? Dazu gibt es viele Studien, die zeigen, dass das Koffein lebensverlängernd ist. Allerdings nur bei Menschen, die höchstens fünf Tassen pro Tag trinken. Andererseits treibt Kaffee den Blutdruck in die Höhe - da sind sich die Experten einig und ist vielleicht für schwere Krankheiten, wie Krebs, verantwortlich. Außerdem gibt es auch die Gefahr der Abhängigkeit vom Koffein: die Kaffeesucht, so sagen die Einen. Die Anderen meinen aber, dass man Kaffee nicht mit Drogen vergleichen kann. Macht Kaffee nun süchtig oder nicht? Ist er ein Muntermacher oder nicht? Leben wir durch Kaffeegenuss länger oder können wir Krebs bekommen?

All diese Unklarheiten sollten uns nicht beunruhigen, denn eines ist sicher: Die Wirkung von Kaffee hängt immer auch von uns selbst ab. Genießen wir unseren Kaffee in Maßen! Und wenn wir fest daran glauben, dass uns ein Espresso am Morgen munter und aktiver macht, dann wird es schon klappen.

(Loumiotis & Näfken, 2016:70)

Word length = 244 words

Text 8 (Goethe-Zertifikat)

Mareike May

An: Annegret

Liebe Annegret,

sicherlich hast du dich schon gefragt, warum ich mich so lange nicht gemeldet habe. Nun, ich bin seit zwei Monaten in Paris und mache hier mein Auslandspraktikum. Meine Gastfamilie, Familie Dupont, ist sehr nett. Sie hat mich vom Flughafen mit dem Taxi abgeholt und wir haben uns sofort gut verstanden. Herr und Frau Dupont sind Lehrer und sprechen sogar ein bisschen Deutsch. Sie möchten, dass ich mit ihnen nur Französisch, mit den beiden Kindern aber nur Deutsch spreche. Am ersten Wochenende habe ich mir Paris allein angesehen, da die Duponts nicht mitkommen konnten. Sie haben kein Auto und außerdem waren die Kinder krank. Mein Praktikum mache ich in einem Krankenhaus, in der Kardiologie. Mein Chef ist mir eher unsympathisch, aber der Assistenzarzt, Alain, ist ganz anders. Er sieht auch sehr gut aus!

Leider ist das Krankenhaus sehr weit von hier entfernt. Es ist nicht besonders groß, aber modern. Das ist super!

Nächste Woche ist Nationalfeiertag und da haben wir drei Tage frei. Da möchte ich ans Mittelmeer fahren. Vielleicht fliege ich auch, dann verliere ich nicht so viel Zeit.

Ich habe noch eine Bitte an dich: Ich habe es nicht geschafft, mich von Frau Sasse, unserer Nachbarin, zu verabschieden. Könntest du vielleicht mal bei

ihr vorbeigehen und ihr Bescheid sagen, wo ich bin? Sie ist ja schon so alt und macht sich vielleicht Sorgen, warum ich sie nicht mehr besuche. So, das wär's für heute!

Sei lieb begrüßt

von Mareike

(Loumiotis & Näfken, 2016:74)

Word length = 246 words

Here follows the table presenting all the information regarding the word length of the twelve texts considered:

Table 3: Length (Goethe-Zertifikat A2)

Goethe-Zertifikat A2 Modellsatz Erwachsene			
Text	Genre	Topic	Length (words)
Text 1	newspaper article	TV show	182
Text 2	e-mail	Life and study in Hamburg	257
Goethe-Zertifikat A2 Modellsatz Jugendliche			
Text	Genre	Topic	Length
Text 3	newspaper article	Youth Newspaper	189
Text 4	e-mail	Invitation to one's home	229
Mit Erfolg zum Goethe Zertifikat A2			
Text	Genre	Topic	Length
Text 5	newspaper article	Example of integration	199

Text 6	e-mail	Class reunion	251
Text 7	newspaper article	Effects of coffee	244
Text 8	e-mail	Internship abroad	246
Text 9	newspaper article	Wild birds	236
Text 10	e-mail	Problems with girlfriend	226
Text 11	newspaper article	Remote work	261
Text 12	e-mail	Accident	253

The word count values for the six newspaper articles range from 182 to 261 and their average is 218.5 with a variance of 943.767, whereas for the six emails, the word count values vary from 226 to 257 and their average stands at 246.667 with a variance of 170.267. Hence, I will proceed as follows: first, I will provide ChatGPT with a prompt without specifying the word length, as acquirers who want to read texts on ChatGPT might have little idea of the appropriate word length for each level of proficiency. Then, I will provide ChatGPT with a prompt specifying the word length of the text to be generated. It would be sensible to aim for a text length of approximately 220 words for newspaper articles and around 250 words for emails when assessing ChatGPT's ability to generate texts at the A2 level. By generating texts of this length, it will be possible to effectively evaluate the model's accuracy in producing content suitable for A2-level readers. If ChatGPT consistently generates texts of this length, it would indicate a high level of accuracy in this aspect. Conversely, if it appears not to be able to produce texts of the desired length, it may suggest limitations in its accuracy.

3.2. Readability

Secondly, it is paramount to analyze the readability of texts generated by ChatGPT. Indeed, comprehensibility stands as one of the cornerstones of the input, it appears to be reasonable to consider the readability of the texts that language acquirers should be exposed to. Traditionally, two indices are employed to assess readability, namely the Flesch Kincaid index and the Gunning Fog Index. However, these indices, designed for English texts, may not be suitable for German due to linguistic differences, particularly in word syllable counts. Therefore, I will adopt the Wiener Sachtextformel developed by Richard Bamberger and Erich Vanecek (1984). This formula considers several factors, including the percentage of words with three or more syllables, average sentence length, percentage of words with more than six letters, and percentage of monosyllabic words. Since that this formula has undergone four revisions, I will rely on the latest version (2006):

MS = the percentage of the words with three syllables or more

SL = average length of sentences

$$WSTF = 0,2744*MS + 0,2656*SL - 1,693$$

(Bamberger, 2006, cited in Rottensteiner, 2010:3897)

The resulting scale corresponds to a school level ranging from 4 to 15. In order to determine the range for A2-level texts, I have analyzed the same twelve texts

mentioned before, and have calculated the WSTF for each text. In this context, it is important to specify that, when it comes to the number of sentences, I have chosen to define them as strings of words ending with a full stop, an exclamation mark, or a question mark. Therefore, strings of words with a semicolon will not be considered sentences within this analysis, as illustrated in the following examples:

Text 9 (Goethe-Zertifikat)

Ein Tierarzt erzählt

Wilde Vögel in der Klinik

Mario Walters ist Tierarzt, aber ein ganz besonderer: Er kümmert sich um große Vögel, wie Eulen, Adler und Falken und hat jeden Tag mit ihnen zu tun. So kann es passieren, dass sie ihn manchmal verletzen. Aber das findet er nicht schlimm. Wichtiger ist für ihn, dass er diesen Vögeln hilft, denn oft sind sie krank. Meist sind ihre Flügel gebrochen oder sie sind fast verhungert.

Die meisten der Vögel verletzen sich, wenn sie gegen fahrende Autos oder Fensterscheiben fliegen. Glücklicherweise gehen die Stromunfälle in den letzten Jahren zurück. Konstant bleibt aber die Zahl der Tiere, die jedes Jahr in Walters' Pflegestation sind so zwischen 60 und 80 Tieren.

Das Wichtigste ist für uns, dass sich die Tiere nicht an uns gewöhnen. Wenn die Vögel wieder gesund sind, sollen sie wieder in der Freiheit leben können", sagt er. Oft bekommt er Tiere, die privat gehalten wurden. Vor zwei

Jahren übergab ihm die Landespolizei 20 Tiere aus Privathaltung, auch seltene Schneeeulen und sogar einen Uhu. Walters ist wütend: „Diese Tiere gehören in die Wildnis und nicht in Käfige. Kein gesunder Vogel gehört in einen Käfig. Auch keine Wellensittiche oder Kanarienvögel. Um vor allem Kinder zu informieren, machen wir in den großen Ferien sowie zu Weihnachten und Ostern Workshops für Schüler. Eltern können auch kommen. Letztes Jahr hatten wir 2000 Besucher und in diesem Jahr erwarten wir noch mehr.

(Loumiotis & Näfken, 2016:86)

Sentences = 19

Text 10 (Goethe-Zertifikat)

Hallo Patrick,

ich melde mich schon wieder, da ich deine Hilfe brauche. Wie ich dir gestern geschrieben habe, habe ich Probleme mit Sandra. Morgen sind wir schon drei Jahre zusammen und ich möchte gern in eine gemeinsame Wohnung ziehen. Aber Sandra möchte das nicht. Sie will bei ihrer Mutter wohnen bleiben, weil das billiger ist. Ich habe ihr gesagt, dass ich genug Geld für eine gemeinsame Wohnung verdiene. Die Mieten sind ja hier nicht so hoch. Aber Sandra will erst ihr Studium beenden und eine Arbeit finden. Wir hätten uns fast gestritten, aber sie ist ja ein ruhiger Mensch und streitet nie.

Ich wollte dieses Jahr endlich mal meine Traumreise machen und nach Südafrika fahren. Das will Sandra auch nicht, weil ich dann alles bezahle. Sie möchte eine so große Reise erst machen, wenn sie selbst Geld verdient. Ich möchte aber nicht mehr warten, denn wir werden ja nicht jünger. Und wer weiß, ob Sandra in einem Jahr wirklich mit dem Studium fertig ist. Sie meint, ich soll die Reise allein machen. Aber so eine Reise ohne sie, das gefällt mir nicht. Was soll ich allein im Urlaub machen?

Jetzt haben wir uns geeinigt, dass wir im August ein paar Tage nach Florenz fahren. Das ist nicht so weit und auch nicht so teuer. Lust habe ich nicht, aber was soll ich machen?

Liebe Grüße

Paul

(Loumiotis & Näfken, 2016:90)

Sentences = 21

Additionally, for newspaper articles, I have counted the title and the subtitles as one sentence each, even though it does not end with a full stop (such as “Ein Tierarzt erzählt” and “Wilde Vögel in der Klinik” in the text 9, which are considered as two sentences). For emails, I have not counted the strings of words introducing the addressee and the addresser, but have focused solely on the content of the email. Furthermore, I considered the final greeting as one single sentence (such as “Liebe Grüße Paul” in the text 10, which has been regarded as one single sentence).

Here follows the table presenting all the information regarding the WSTF value of the twelve texts considered:

Table 4: Readability (Goethe-Zertifikat A2)

Goethe-Zertifikat A2 Modellsatz Erwachsene						
Text	Genre	Topic	Length (words)	Sentences	Words with three syllables or more	WSTF
Text 1	newspaper article	TV show	182	16	24	4.947
Text 2	e-mail	Life and study in Hamburg	257	25	29	4.134
Goethe-Zertifikat A2 Modellsatz Jugendliche						
Text	Genre	Topic	Length	Sentences	Words with three syllables or more	WSTF
Text 3	newspaper article	Youth Newspaper	189	17	31	5.761
Text 4	e-mail	Invitation to one's home	229	23	24	3.827
Mit Erfolg zum Goethe Zertifikat A2						
Text	Genre	Topic	Length	Sentences	Words with three syllables or more	WSTF
Text 5	newspaper article	Example of integration	199	12	47	9.192
Text 6	e-mail	Class reunion	251	24	35	4.911
Text 7	newspaper article	Effects of coffee	244	19	33	5.429
Text 8	e-mail	Internship abroad	246	22	30	4.623
Text 9	newspaper article	Wild birds	236	19	29	4.978
Text	e-mail	Problems with	226	21	14	2.865

10		girlfriend				
Text 11	newspaper article	Remote work	261	24	42	5.611
Text 12	e-mail	Accident	253	25	26	3.815

The values of the six newspaper articles vary from 4.947 to 9.192 and their average is 5.986 with a variance of 2.575, whereas for the six e-mails, the values range from 2.865 to 4.911 and their average amounts to 4.029 with a variance of 0.517. Hence, although there is no direct correlation between the WSTF classification and the CEFR, it specifies for which school level an informational text is suitable and therefore it seems to be reasonable to assume that newspapers with a value of 6 and e-mails with a value of 4 align with level A2 proficiency. Therefore, if ChatGPT can produce texts within this range, it signifies that it accurately meets this criterion; otherwise, it indicates inaccuracies in ChatGPT's output in this context.

3.3.Vocabulary

The third criterion focuses on vocabulary, specifically evaluating whether the vocabulary employed in the generated texts corresponds to the proficiency levels delineated in the Common European Framework of Reference for Languages. Also in this context, the materials provided by the Goethe Institut for each CEFR level can serve as a reference point to gauge the accuracy of ChatGPT in generating texts at a specific proficiency level. Consequently, the A2 word list will be used as the benchmark for evaluating the vocabulary proficiency

of ChatGPT-generated texts. By cross-referencing the words used in the generated texts with those listed in the A1 and A2 word lists (Goethe-Institut, 2013), we can ascertain whether ChatGPT can accurately produce texts suitable for learners at the A2 proficiency level. However, it must be pointed out that articles, pronouns, proper nouns, and numbers are considered as known words, even though they are not present in the A1 and A2 word lists. Hence, they will not be counted among the unknown words in this evaluation. In the following example, the bold words represent the unknown ones:

Text 11 (Goethe-Zertifikat)

Wohnen und Leben Arbeiten zu Hause

Homeoffice heißt das neue **Zauberwort**. Das bedeutet, dass man seinen Arbeitsplatz in den eigenen vier **Wänden** hat. Immer mehr Menschen arbeiten so. Dabei gefällt es den meisten, dass lange **Fahrten** zum Büro und wieder zurück. vorbei sind, dass man sich nicht in die volle Bahn **drängeln** muss, dass man sich nicht mit langweiligen Kollegen unterhalten muss. Schon über elf Prozent der **Berufstätigen** Deutschlands arbeiten ab und zu oder **ständig** so. Doch kann das **funktionieren**? Möchte man nicht vielleicht doch noch schnell mal die Wäsche **bügeln**, einkaufen gehen, die Freundin oder den Freund anrufen? Nur wer **diszipliniert** ist und eine gute **Arbeitsatmosphäre** schafft, wird bei so einem Arbeitsplatz glücklich. Man braucht eine schnelle und gute **Internetverbindung** **sowie** den richtigen

Schreibtisch. Wo der steht, ist egal: im Wohnzimmer, im **Gästezimmer** oder im Schlafzimmer.

Wichtig ist die **Höhe** des **Schreibtisches**, die von der Größe des **arbeitenden** Menschen **abhängt**, und der richtige Stuhl. Man sollte auch nicht so viel sitzen, **sondern** manchmal aufstehen. Das ist für das **Herz** und die **Figur** gut und erlaubt auch ein **konzentriertes** Arbeiten. Kleiner **Trick**: Der Drucker steht in einem **anderen** Zimmer. So muss man aufstehen, um ihn zu erreichen. Außerdem sollte der **Schreibtisch möglichst** leer sein. Bücher, Papiere, **Multifunktionsgeräte usw.** können in **Regalen** oder Sideboards stehen oder liegen. Aber auch das Licht spielt eine große **Rolle**. **Sonnenlicht** sollte **seitlich** vom Fenster **einfallen**. **Optimal** sind auch Licht von oben und eine **Schreibtischlampe**. Und: Weniger ist mehr. Drei Stifte **genügen** und vielleicht noch ein Foto der Lieben oder ein lustiger Kalender. Mehr braucht man nicht.

(Loumiotis & Näfken, 2016:102)

Unknown words = 36

Here are the results of such cross-referencing:

Table 5: Vocabulary (Goethe-Zertifikat A2)

Goethe-Zertifikat A2 Modellsatz Erwachsene					
Text	Genre	Topic	Length (words)	Unknown words	Percentage of known words
Text 1	newspaper article	TV show	182	22	87,912%
Text 2	e-mail	Life and study in Hamburg	257	27	89,494%
Goethe-Zertifikat A2 Modellsatz Jugendliche					
Text	Genre	Topic	Length	Unknown words	Percentage of known words
Text 3	newspaper article	Youth Newspaper	189	22	88,360%
Text 4	e-mail	Invitation to one's home	229	20	91,266%
Mit Erfolg zum Goethe Zertifikat A2					
Text	Genre	Topic	Length	Unknown words	Percentage of known words
Text 5	newspaper article	Example of integration	199	51	74,372%
Text 6	e-mail	Class reunion	251	36	85,657%
Text 7	newspaper article	Effects of coffee	244	54	77,869%
Text 8	e-mail	Internship abroad	246	24	90,244%
Text 9	newspaper article	Wild birds	236	43	81,780%
Text 10	e-mail	Problems with girlfriend	226	9	96,617%
Text 11	newspaper article	Remote work	261	36	86,207%

Text 12	e-mail	Accident	253	32	87,352%
---------	--------	----------	-----	----	---------

The percentages of known words for newspaper articles range from 74.372% to 88.36%, their average is 82.75 % with a variance of 33.005, while for emails, the percentages vary from 87.352% to 96.617%, their average stands at 90.105% with a variance of 13.177. Therefore, if approximately 83% of the words in the generated newspaper articles and 90% of the words in the emails align with those listed in the A1 and A2 word list, we can conclude that ChatGPT accurately meets this criterion. Conversely, if there are more unknown words in the generated texts than expected, it indicates inaccuracies in ChatGPT's output.

3.4. Conclusion

It has been explained how I shall examine the possibility of generating texts using ChatGPT to provide language acquirers with comprehensible input, specifically targeting proficiency level A2 of the Common European Framework of Reference for Languages (CEFR) in the German language.

Firstly, the length of the generated texts will be examined to determine the extent to which it can be controlled or determined. This criterion will be assessed by providing ChatGPT with prompts both with and without specifying word length. As has been pointed out, I will aim for approximately 220 words for newspaper articles and around 250 words for emails. Consistency in generating texts of these lengths would indicate a high level of accuracy, while deviation might suggest limitations.

Secondly, text readability will be evaluated using the Wiener Sachtextformel (WSTF) developed by Bamberger and Vanecek. Although WSTF classification does not directly correlate with CEFR levels, it has been explained that it is reasonable to assume that newspapers with a WSTF value of 6 and emails with a value of 4 align with A2 proficiency. Thus, if ChatGPT produces texts within this range, it will meet this criterion; otherwise, it suggests inaccuracies.

Finally, the vocabulary in generated texts will be assessed by cross-referencing words in the generated texts and words used with A1 and A2 word lists. If approximately 83% of words in newspaper articles and 90% in emails align with these lists, ChatGPT can be considered accurate in this regard. Conversely, a significant number of unknown words implies inaccuracies.

In conclusion, with regard to the prompt provided to ChatGPT, I will employ zero-shot prompting, a technique enabling the model to make predictions on unseen data without further training. This approach appears to be particularly useful in assessing how effectively ChatGPT can generate texts with minimal input. As mentioned earlier, I will provide ChatGPT with two different prompts. In the first prompt, I will specify the text type (newspaper article or email), proficiency level, and the topic of the text. In the second prompt, I will also add the desired text length. Therefore, I will analyze a total of 24 texts (12 from the first prompt and 12 from the second). If the aforementioned criteria are met, then it will be possible to assess ChatGPT's accuracy in generating comprehensible output, whereas failure to meet these criteria will point out certain inaccuracies in the generation of comprehensible input.

4. Experiment and Evaluation

This section delves into the experimentation involving texts generated by ChatGPT and the subsequent evaluation of such texts. As previously outlined, a set of 24 texts has been generated, comprising 12 texts without controlled length and 12 texts with controlled length. When it comes for texts without controlled length, the following prompt has been used.

Table 6: Prompts

	For newspaper articles	For e-mails
Texts without controlled length	Create a newspaper article in German about [topic]. The text level should be A2 according to CEFR.	Create an e-mail in German about [topic]. The text level should be A2 according to CEFR.
Texts with controlled length	Create a newspaper article in German about [topic]. The length of the text should be 220 words. The text level should be A2 according to CEFR.	Create an e-mail in German about [topic]. The length of the text should be 250 words. The text level should be A2 according to CEFR.

These texts deal with the same topics as the 12 texts from the Goethe-Zertifikat A2, as detailed in the table below:

Table 7: Texts (ChatGPT)

Text	Controlled length	Genre	Topic
Text 1	no	newspaper article	TV show
Text 2	yes	newspaper article	TV show
Text 3	no	e-mail	Life and study in Hamburg
Text 4	yes	e-mail	Life and study in Hamburg
Text 5	no	newspaper article	Youth Newspaper
Text 6	yes	newspaper article	Youth Newspaper
Text 7	no	e-mail	Invitation to one's home
Text 8	yes	e-mail	Invitation to one's home
Text 9	no	newspaper article	Example of integration
Text 10	yes	newspaper article	Example of integration
Text 11	no	e-mail	Class reunion
Text 12	yes	e-mail	Class reunion
Text 13	no	newspaper article	Effects of coffee
Text 14	yes	newspaper article	Effects of coffee
Text 15	no	e-mail	Internship abroad
Text 16	yes	e-mail	Internship abroad
Text 17	no	newspaper article	Wild birds
Text 18	yes	newspaper article	Wild birds

Text 19	no	e-mail	Problems with girlfriend
Text 20	yes	e-mail	Problems with girlfriend
Text 21	no	newspaper article	Remote work
Text 22	yes	newspaper article	Remote work
Text 23	no	e-mail	Accident
Text 24	yes	e-mail	Accident

Before delving into the experiment using these texts, it is crucial to point out some limitations of the texts generated by ChatGPT, in particular with regard to the e-mails. Indeed, I had to designate names for both the addressee and the addresser, as ChatGPT would otherwise leave these sections blank. Consequently, I opted for the names “Marco” and “Anna” respectively. However, some texts exhibit limitations. For instance, in text 7, ChatGPT could not understand whether “Marco” is a masculine or feminine name and simply wrote “Liebe/r Marco,” although the correct option would be “Lieber Marco”: Similarly, in text 24, ChatGPT even assumed a feminine gender for Marco, resulting in “Liebe Marco”: Furthermore, despite specifying the names of the addressee and the addresser, some blanks remained, as observed in text 8, where ChatGPT omitted the address, leaving “[Deine Adresse hier einfügen]”. This pattern repeats in text 12, where date, time, and place are denoted as “[Datum],” “[Uhrzeit],” and “[Ort],” and in text 16, with placeholders “[Unternehmen/Organisation],” “[Nachname],” and “[Position].” Additionally, ChatGPT occasionally lacked creativity, as can be observed in text 15's "Firma XYZ" instead of a more fitting company name. These

limitations are likely to derive from insufficient information provided to ChatGPT, resulting in placeholders for name, address, date, time, and place. Despite these limitations, ChatGPT generally succeeded in creating texts that adhered to specified topics. This capability is illustrated, among others, in text 4:

Text 4 (ChatGPT)

Betreff: Leben und Studieren in Hamburg

Lieber Marco,

ich hoffe, diese E-Mail findet dich gut! Mein Name ist Anna, und ich studiere zurzeit in Hamburg. Ich wollte dir ein bisschen über das Leben und das Studium hier erzählen.

Hamburg ist eine wunderschöne Stadt mit vielen Möglichkeiten. Es gibt hier viele Parks, wo man spazieren gehen oder sich einfach entspannen kann. Auch das Nachtleben ist sehr lebendig, mit vielen Bars und Clubs für jeden Geschmack.

Das Studium hier ist auch sehr interessant. Ich studiere Wirtschaftswissenschaften an der Universität Hamburg. Die Vorlesungen sind auf Deutsch, aber es gibt auch viele Kurse, die auf Englisch angeboten werden. Die Professoren sind sehr hilfsbereit und stehen immer zur Verfügung, um Fragen zu beantworten.

Eine Herausforderung beim Studium in Hamburg ist die Wohnungssuche. Es kann manchmal schwierig sein, eine bezahlbare Wohnung zu finden,

besonders in der Nähe der Universität. Aber mit Geduld und ein wenig Hilfe von Mitbewohnern oder anderen Studierenden findet man oft eine passende Unterkunft.

Insgesamt genieße ich mein Leben und mein Studium hier in Hamburg sehr. Es ist eine lebendige und vielfältige Stadt, die für jeden etwas zu bieten hat. Ich hoffe, diese Informationen helfen dir bei deiner Entscheidung, ob du hier studieren möchtest.

Viele Grüße aus Hamburg,

Anna

Moreover, the above-mentioned issues were absent in the articles, indicating that ChatGPT performs more proficiently in generating this type of texts, as can be seen, among others, in text 2:

Text 2 (ChatGPT)

Faszinierende Unterhaltung auf dem Bildschirm: Die neue TV-Show "Abenteuerlust"

In der aufregenden Welt des Fernsehens gibt es eine neue Show, die die Herzen der Zuschauer erobert: "Abenteuerlust". Die Show, die auf einem der führenden deutschen Sender ausgestrahlt wird, bietet ein spannendes und unterhaltsames Erlebnis für die ganze Familie.

In "Abenteuerlust" nehmen mutige Teilnehmer an verschiedenen Abenteuern und Herausforderungen teil. Von waghalsigen Klettertouren über spannende

Rätsel bis hin zu aufregenden Survival-Situationen ist für jeden etwas dabei. Die Teilnehmer müssen ihr Geschick, ihre Intelligenz und ihren Teamgeist unter Beweis stellen, um die Herausforderungen zu meistern und den Sieg zu erringen.

Die Zuschauer können sich auf packende Wettkämpfe und überraschende Wendungen freuen, während sie die Teilnehmer auf ihrer Reise begleiten. Die Moderatorin der Show, die charismatische Julia Müller, führt die Zuschauer durch die einzelnen Episoden und sorgt für jede Menge Spaß und Spannung.

"Abenteuerlust" ist nicht nur eine TV-Show, sondern auch eine Quelle der Inspiration für Abenteurer jeden Alters. Die Show ermutigt die Zuschauer, ihre eigenen Grenzen zu überwinden und neue Herausforderungen anzunehmen.

Verpassen Sie nicht die nächste Folge von "Abenteuerlust" und tauchen Sie ein in die aufregende Welt des Abenteuers!

After highlighting the limitations in texts generated by ChatGPT, it is now possible to consider the three specified criteria: length, readability, and vocabulary. First of all, the texts generated without controlled length will be examined; then the texts with controlled length will be taken into account.

4.1. Texts without controlled length

In this part, texts generated by ChatGPT without specified length controls will be analyzed. This approach will allow us to determine if ChatGPT can naturally produce texts of an appropriate length for learners at the A2 proficiency level in German, without the need to predefine text lengths.

4.1.1. Length

First of all, word length will be taken into account. The outcomes of this experiment are detailed below:

Table 8: Length (ChatGPT without controlled length)

Text	Genre	Topic	Length (words)
Text 1	newspaper article	TV show	207
Text 3	e-mail	Life and study in Hamburg	188
Text 5	newspaper article	Youth Newspaper	207
Text 7	e-mail	Invitation to one's home	111
Text 9	newspaper article	Example of integration	284
Text 11	e-mail	Class reunion	124
Text 13	newspaper article	Effects of coffee	248
Text 15	e-mail	Internship abroad	150

Text 17	newspaper article	Wild birds	230
Text 19	e-mail	Problems with girlfriend	123
Text 21	newspaper article	Remote work	173
Text 23	e-mail	Accident	135

Hence, word count values of the six articles without controlled length range from 173 to 284 and their average is 224.833 with a variance of 1477.367, whereas the word count values of the six e-mails without controlled length range from 111 to 183 and their average is 138.5 with a variance of 760.3.

It can be observed that the average word count for articles (224.833) is close to the previously specified target of 220 words. However, the substantial variance (1477.367) suggests that ChatGPT may still have some limitations in consistently achieving this target. Additionally, the average word count for emails (183.5) falls significantly short of the target of 250 words, highlighting further inaccuracies. Hence it is possible to conclude that ChatGPT has significant limitations in generating texts suitable for the A2 level of proficiency when the length of the texts is not specified

4.1.2. Readability

When it comes to the readability of the texts generated by ChatGPT, it has been explained that if ChatGPT can achieve a WSTF score of 6 for newspapers and 4 for emails, it indicates it effectively meets this standard. Here follow the results of the experiment:

Table 9: Readability (ChatGPT without controlled length)

Text	Genre	Topic	Length (words)	Sentences	Words with three syllables or more	WSTF
Text 1	newspaper article	TV show	207	13	51	9.297
Text 3	e-mail	Life and study in Hamburg	188	15	34	6.598
Text 5	newspaper article	Youth Newspaper	207	13	51	9.297
Text 7	e-mail	Invitation to one's home	111	9	13	4.796
Text 9	newspaper article	Example of integration	284	18	67	8.971
Text 11	e-mail	Class reunion	124	10	13	4.477
Text 13	newspaper article	Effects of coffee	248	17	58	8.599
Text 15	e-mail	Internship abroad	150	10	35	8,694
Text 17	newspaper article	Wild birds	230	19	44	6.772
Text 19	e-mail	Problems with girlfriend	123	10	9	3.582
Text 21	newspaper article	Remote work	173	14	45	8.727
Text 23	e-mail	Accident	135	11	17	5.022

The values of the six articles without controlled length range from 6.772 to 9.297 and their average is 8.611 with a variance of 0.886, whereas the values for the six e-mails without controlled length range from 3.582 to 8.694 and their average is 5.528 with a variance of 3.370.

The analysis of readability scores for texts generated by ChatGPT without controlled length presents notable challenges in text complexity management. For articles, the average readability score of those without controlled length stands at 8.611, considerably higher than the target of 6. In the case of emails, the results are somewhat more encouraging but still not fully aligned with the targets. Uncontrolled emails have an average readability score of 5.528, slightly above the desired mark of 4. Despite the emails being nearer to the target, the overall results demonstrate limitations in producing texts that are adequately simplified for beginner learners.

4.1.3. Vocabulary

With regard to vocabulary, it has been established that if approximately 83% of the words in the generated newspaper articles and 90% of the words in the emails correspond with those listed in the A1 and A2 word list, it is possible to assert that ChatGPT effectively meets this standard. Below are the experiment outcomes:

Table 10: Vocabulary (ChatGPT without controlled length)

Text	Genre	Topic	Length (words)	Unknown words	Percentage of known words
Text 1	newspaper article	TV show	207	31	85.024%
Text 3	e-mail	Life and study in Hamburg	188	74	60.638%
Text 5	newspaper article	Youth Newspaper	207	31	85.024%
Text 7	e-mail	Invitation to one's home	111	46	58.559%
Text 9	newspaper article	Example of integration	284	29	89.789%
Text 11	e-mail	Class reunion	124	54	56.452%
Text 13	newspaper article	Effects of coffee	248	41	83.468%
Text 15	e-mail	Internship abroad	150	50	66.667%
Text 17	newspaper article	Wild birds	230	27	88.261%
Text 19	e-mail	Problems with girlfriend	123	49	60.163%
Text 21	newspaper article	Remote work	173	29	83.237%
Text 23	e-mail	Accident	135	66	51.111%

The percentages of known words for the six articles without controlled length range from 83.237% to 89.789% and their average is 85.801% with a variance of 7.036, whereas the percentages for the six e-mails without controlled length range from 51,111% to 66,667% and their average is 58,932% with a variance of 26.343.

When evaluating the vocabulary adequacy in articles generated by ChatGPT, it appears that the system performs relatively well in aligning with the A2 proficiency level, as the average percentage of known words in articles without length control is 85.801%. This suggests that ChatGPT is capable of producing articles that largely meet the vocabulary requirements for an A2 level, achieving approximately 83% alignment with the target. However, the performance is less satisfactory in the context of e-mails. The average known word percentage in e-mails without length control stands at only 58.932%, significantly below the target of 90%. This discrepancy highlights substantial limitations in ChatGPT's ability to consistently apply A2-level vocabulary in e-mails, indicating a need for further refinement in this area.

4.2. Texts with controlled length

In this part, texts generated by ChatGPT with specified length controls will be analyzed. This analysis will help determine whether specifying word length leads to improved results across the three criteria: length, readability, and vocabulary.

4.2.1. Length

First of all, word length will be taken into account. In accordance with the methodology outlined in the third section, I will aim at generating articles comprising approximately 220 words for newspapers and around 250 words for

emails. The outcomes of this experiment are detailed below:

Table 11: Length (ChatGPT with controlled length)

Text	Genre	Topic	Length (words)
Text 2	newspaper article	TV show	188
Text 4	e-mail	Life and study in Hamburg	206
Text 6	newspaper article	Youth Newspaper	185
Text 8	e-mail	Invitation to one's home	180
Text 10	newspaper article	Example of integration	199
Text 12	e-mail	Class reunion	165
Text 14	newspaper article	Effects of coffee	172
Text 16	e-mail	Internship abroad	191
Text 18	newspaper article	Wild birds	178
Text 20	e-mail	Problems with girlfriend	189
Text 22	newspaper article	Remote work	199
Text 24	e-mail	Accident	182

Word count values of the six articles with controlled length vary from 172 to 199 and their average is 186.833 with a variance of 119.767, while the word count values of the six e-mails with controlled length vary from 165 to 206 and their average is 185.5 with a variance of 185.1.

Thus, the analysis of texts generated by ChatGPT with controlled length settings reveals intriguing results. When compared to the average word count of 224.833 for articles generated without length control, the controlled articles significantly dropped to 186.833. Surprisingly, this is well below the target of 250 words, showcasing a counterintuitive outcome where texts with specified length constraints were less accurate than those generated without such controls. As for the emails, though the length-controlled ones approached the intended 220-word target more closely than their uncontrolled counterparts, they still fell short, highlighting inaccuracies in achieving the specified word count. These findings suggest that ChatGPT may have limitations in consistently managing text length effectively.

4.2.2. Readability

When it comes to the readability of the texts generated by ChatGPT, it has been explained that if ChatGPT can achieve a WSTF score of 6 for newspapers and 4 for emails, it indicates it effectively meets this standard. Here follow the results of the experiment:

Table 12: Readability (ChatGPT with controlled length)

Text	Genre	Topic	Length (words)	Sentences	Words with three syllables or more	WSTF
Text 2	newspaper article	TV show	188	11	56	11.02

Text 4	e-mail	Life and study in Hamburg	206	17	35	6.188
Text 6	newspaper article	Youth Newspaper	185	13	46	8.910
Text 8	e-mail	Invitation to one's home	180	13	23	5.491
Text 10	newspaper article	Example of integration	199	14	49	8.839
Text 12	e-mail	Class reunion	165	12	18	4,952
Text 14	newspaper article	Effects of coffee	172	15	44	8.372
Text 16	e-mail	Internship abroad	191	14	51	9.257
Text 18	newspaper article	Wild birds	178	13	39	7.956
Text 20	e-mail	Problems with girlfriend	189	12	18	5.104
Text 22	newspaper article	Remote work	199	14	56	9.804
Text 24	e-mail	Accident	182	14	25	5.529

The values of the six articles with controlled length vary from 7.956 to 11.02 and their average is 9.150 with a variance of 1.222, while the values for the six e-mails with controlled length vary from 4.952 to 9.257 and their average is 6.087 with a variance of 2.596.

The analysis of readability scores reveals considerable challenges in managing text complexity. Notably, the readability score for articles with controlled length increases to 9.150, up from 8.611 for those without controlled length, suggesting that the controlled articles are even more challenging for readers. This result is counterintuitive, as one might expect controlled lengths to aid in simplifying text

complexity. Similarly, in the case of emails, the results are not much improved. The average readability score for uncontrolled emails stands at 5.528, but for those with controlled length, it worsens slightly to 6.087. This increase indicates that even the controlled emails may still be too complex for German learners at the A2 level, failing to achieve a WSTF score specified in the previous section. These findings underscore a significant issue: both articles and emails with controlled lengths tend to be more challenging than their uncontrolled counterparts, highlighting a clear limitation in producing texts that are adequately simplified for beginner learners.

4.2.3. Vocabulary

With regard to vocabulary, it is now possible to assess whether texts with controlled length exhibit better vocabulary alignment with the target language level compared to those without controlled length. This comparison will help determine if specifying text length contributes to more effective vocabulary usage suitable for the intended proficiency level. Below are the experiment outcomes:

Table 13: Vocabulary (ChatGPT with controlled length)

Text	Genre	Topic	Length (words)	Unknown words	Percentage of known words
Text 2	newspaper article	TV show	188	21	88.830%
Text 4	e-mail	Life and study in Hamburg	206	70	66.019%

Text 6	newspaper article	Youth Newspaper	185	16	91.351%
Text 8	e-mail	Invitation to one's home	180	58	67.778%
Text 10	newspaper article	Example of integration	199	23	88.442%
Text 12	e-mail	Class reunion	165	63	61.818%
Text 14	newspaper article	Effects of coffee	172	29	82.140%
Text 16	e-mail	Internship abroad	191	71	62.827%
Text 18	newspaper article	Wild birds	178	15	91.573%
Text 20	e-mail	Problems with girlfriend	189	57	69.841%
Text 22	newspaper article	Remote work	199	27	86.432%
Text 24	e-mail	Accident	182	49	73.077%

The percentages of known words for the six articles with controlled length vary from 82.14% to 91.573% and their average is 88.128% with a variance of 12.317, while the percentages for the six e-mails with controlled length vary from 61.818% to 73.077% and their average is 66.893% with a variance of 26.344.

In the case of vocabulary adequacy, the analysis of articles generated by ChatGPT shows that the system performs reasonably well in aligning with the A2 proficiency level. When comparing the average percentage of known words, controlled articles show an increase from 85.801% in uncontrolled articles to 88.128% in articles with length control. This improvement suggests that ChatGPT is capable of producing articles that meet the vocabulary requirements for the A2 level more effectively

when text length is specified, achieving approximately 83% alignment with the target. However, the performance in the context of emails is less satisfactory. Although there is an improvement in vocabulary adequacy in emails with controlled length (from 58.932% in uncontrolled emails to 66.893%) both figures remain significantly below the desired target of 90%. This substantial gap underscores limitations in ChatGPT's ability to consistently apply A2-level vocabulary in emails, highlighting a clear need for further refinement to enhance its effectiveness in this specific application.

4.3. Conclusion

In conclusion, the evaluation of the experiment carried out in this section indicates that ChatGPT still shows considerable limitations in generating texts that provide comprehensible input for foreign language acquirers. This is evident across three critical areas of analysis: length, readability, and vocabulary adequacy.

Firstly, length control in text generation demonstrates inconsistent results. Indeed, as for articles, controlled texts are even less accurate in meeting length specifications than those generated without such constraints. On the other hand, emails achieve a closer but still off-target average of word length. Hence, these findings underline a fundamental challenge in ChatGPT's ability to manage text length effectively.

Secondly, the analysis of readability scores reveals substantial difficulties in maintaining appropriate text complexity. It has been pointed out that texts with length control tend to have a higher readability score and thus appear to be even

more challenging to read. For emails, although closer to the desired standard, the readability levels suggest that the texts may still be too complex for A2 level German learners, pointing to significant limitations in generating suitably simplified content.

Lastly, while vocabulary alignment in articles shows some promise, achieving approximately the A2 level target, the results for emails are less encouraging, as both uncontrolled and controlled emails are substantially below the target of the known vocabulary in A2 texts. This marked inconsistency highlights critical limitations in ChatGPT's capacity to consistently employ A2-level vocabulary, especially in email texts.

Overall, the outcomes of the experiment carried out demonstrate that while ChatGPT exhibits certain strengths, its capacity to produce linguistically and contextually appropriate texts for A2 level learners is currently limited and requires further enhancement to better serve foreign language acquirers.

5. Conclusion

This thesis has critically evaluated the potential of ChatGPT, a sophisticated language model by OpenAI, in generating German texts that align with Stephen Krashen's Input Hypothesis. As has been explained, according to Krashen, successful language acquisition hinges on the learner's exposure to comprehensible input that is just slightly beyond their current comprehension level. This study specifically aimed to determine if ChatGPT could produce texts appropriate for learners at the A2 level of the Common European Framework of Reference for Languages, focusing on how well these texts met criteria such as text length, readability, and vocabulary appropriateness. The experimental phase of this thesis was designed to assess ChatGPT's performance across these dimensions. It might be assumed that while ChatGPT holds potential in certain areas, its application in language learning, particularly at the A2 level, is constrained by several significant limitations.

In the experiment, ChatGPT's ability to control text length showed some promise, yet the outcomes were inconsistent. The model was more successful in generating texts of a predefined length for structured formats like emails than for more variable formats such as newspaper articles. This variability in performance highlighted a fundamental challenge in using ChatGPT for educational purposes, where precise text length can play a critical role in designing learning materials.

Regarding readability, the texts generated by ChatGPT often did not align with the simplicity required for A2 level learners. This misalignment suggests that while ChatGPT can generate grammatically correct and coherent sentences, it struggles

to adjust the complexity of its output to match the specific needs of beginner language learners. This limitation is particularly problematic because the effectiveness of comprehensible input relies not only on linguistic accuracy but also on the ease with which learners can understand and engage with the material.

Moreover, while the vocabulary alignment in articles indicates some success in meeting the A2 level target, the performance in emails is less satisfactory. Both the uncontrolled and controlled emails consistently fail to reach the vocabulary benchmarks set for A2 level texts, highlighting a notable gap in language proficiency. Hence, although there were instances where the vocabulary was appropriately simplified, these were not consistent enough to conclude that ChatGPT can reliably produce texts that meet the language proficiency requirements of all learners.

These findings from the experiment underscore a critical aspect of employing AI like ChatGPT in educational settings: while the technology can simulate certain types of linguistic interactions and provide useful feedback, it does not yet fully align with the nuanced requirements of language learning frameworks like the CEFR. This misalignment raises concerns about the model's current utility in real-world educational applications, where the accuracy of text generation must meet very specific educational objectives.

In conclusion, the findings of this study suggest that while ChatGPT displays notable capabilities in text generation, its role in facilitating comprehensible input for foreign language acquisition, particularly at the A2 proficiency level, appears to be constrained by several limitations. These preliminary observations indicate a

potential misalignment with the specific needs of language learners, underscoring an area ripe for further inquiry and development. Despite these challenges, ChatGPT offers notable advantages. As previously mentioned, ChatGPT allows users to select the topics of generated texts, potentially increasing engagement by aligning content with students' interests. The process of generating texts is straightforward and time-efficient, providing a practical tool for continuous language practice. Importantly, as ChatGPT is freely available, it enables learners to access fresh educational content throughout their language learning journey without incurring additional costs. Finally, this research has focused on the capabilities of ChatGPT. Future advancements in AI models could lead to improvements that more effectively meet the three criteria examined and thus better serve as comprehensible input for learners.

Hence, as AI technology continues to evolve, it is possible that models like ChatGPT might become integral to language education. Realizing this potential, however, would likely depend on rigorous testing and carefully tailored adaptations to fit educational contexts. This approach might enable AI technologies to not just meet, but potentially expand upon, the pedagogical goals of language learning programs. This study has thus offered preliminary insights and suggests that further exploration is critical. Future investigations might explore more advanced AI training techniques or delve into refining prompt engineering strategies to better harness the educational capabilities of AI-generated content. By continuing such research, it is necessary to gradually uncover how AI might more effectively support the intricate process of language acquisition and make significant contributions to the field of educational technology.

Bibliography

- Abdelhadi, A. (2023). Exploring Chat GPT Potential in Flipped Classrooms for Improving Lexico-grammatical Features in Algerian EFL Learners' Essay Writing. *International Journal of Early Childhood Special Education* 15(04), pp. 121-128.
- Amini-Salehi, E., Hassanipour, S., Bozorgi, A. et al. (2023). *The ability of Chat-GPT to paraphrase texts and reduce plagiarism*. DOI: 10.2196/preprints.53308.
- Bamberger, R. & Vanecek, E. (1984). *Lesen-Verstehen-Lernen-Schreiben. Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Wien: Verlag für Jugend und Volk.
- Bamberger, R. (2006). *Erfolgreiche Leseerziehung. Theorie und Praxis*. München: Domino.
- Benavides, I. (2023). *CHAT GPT: Una reflexión inicial sobre su uso para fines educativos*. National University of Colombia.
- Božić, V. & Poola, I. (2023). *Chat GPT and education*. DOI:10.13140/RG.2.2.18837.40168.
- Constantino, R., S. Y. Lee, K. S. Cho & S. D. Krashen. 1997. Free voluntary reading as a predictor of TOEFL scores. *Applied Language Learning* 8, pp. 111-18.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Dawa, T., Dhendup, S., Tashi, S., Rosso, M. A. (2023). *Technology Literacies and Chat GPT: Perspectives on Emerging Technology Adoption in Academic Writing*. DOI: 10.13140/RG.2.2.31177.83042DeLone.
- Dulay, H. & Burt, M. (1977). Remarks on creativity in language acquisition. In Burt, M., Dulay, H., and Finocchiaro, M. *Viewpoints on English as a Second Language*. New York: Regents, pp. 95-126.
- Gimpel, H., Hall, K., Decker, S. et al. (2023). *Unlocking the power of generative AI models and systems such as GPT-4 and ChatGPT for higher education: A guide*

for students and lecturers. Hohenheim Discussion Papers in Business, Economics and Social Science, Universität Hohenheim.

Goethe-Institut (2013). *Goethe-Zertifikat A2 Fit in Deutsch 2: Prüfungsziele, Testbeschreibung*. Ismaning: Hueber.

Hanane, O. (2023). The Effects of Chat GPT Technology Use on Enhancing ESP Students' Writing Proficiency. The Case of Master One Students at the Faculty of Economic, Commercial, and Management Sciences at Chadli Bendjedid University. *Algerian Scientific Journal Platform* 8(5), pp. 74-82.

Hatch, E. (1979). Apply with caution. *Studies in Second Language Acquisition* 2, pp. 123-143.

Krashen, S. D. (1982). *Principles and Practice in Second Language Acquisition*. Oxford: Pergamon Press.

Krashen, S. D. (2004). The case for narrow reading. *Language Magazine* 3(5), pp. 17-19.

Krashen, S. D. (2011). *Free Voluntary Reading*. Libraries Unlimited.

Krashen, S. D. (1989). *Language Acquisition and Language Education Extensions Applications*. New York: Prentice Hall.

Krashen, S. D. (2004). *The Power of Reading, Second Edition: Insights from the Research*. Libraries Unlimited.

Krashen, S. D. (2017). *Comprehensible and Compelling: The Causes and Effects of Free Voluntary Reading*. Libraries Unlimited.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. Arnaud & H. Béjoint. *Vocabulary and applied linguistics*. London: Macmillan, pp. 126–32.

Loumiotis U. & Näfken, A. (2016). *Mit Erfolg zum Goethe-Zertifikat A2*. Stuttgart: Klett Sprachen GmbH.

Marquez J. (2023) GPT-4: cuándo se lanzará la IA más avanzada de OpenAI, cómo funcionará y todas las novedades. Available at SSRN: <https://www.xataka.com/nuevo/gpt-4-que-cuando-sale-como-funciona-toda-inform>

acion

Mercedes, F., Olite, F. M. D., Del, I. et al. (2023). Artículo de posición Chat GPT: Origen, Evolution, Challenges and Impacts on Education. *Revista Cubana de Educacion Medica Superior* 37(2), pp. 1-23.

Mhlanga, D. (2023). *The Value of Open AI and Chat GPT for the Current Learning Environments and the Potential Future Uses*. Available at SSRN: <https://ssrn.com/abstract=4439267> or <http://dx.doi.org/10.2139/ssrn.4439267>.

Morales-Chan, M. (2023). *Explorando el potencial de Chat GPT: Una clasificación de Prompts efectivos para la enseñanza*. Available at SSRN: <http://hdl.handle.net/123456789/1348>.

Paulston, C. (1972). Structural pattern drills: a classification. *Foreign Language Annals* 4, pp. 187-193.

Rottensteiner, S. (2010). Structure, function and readability of new textbooks in relation to comprehension. *Procedia Social and Behavioral Sciences* 2(2), pp. 3892-3898.

Stokes, J., S. Krashen, & J. Kartchner. (1998). Factors in the acquisition of the present subjunctive in Spanish: The role of reading and study. *ITL: International Journal of Applied Linguistics* 121(1), pp. 19-25.

초록

독일어 습득에 인공지능의 적용

- 이해가능한 입력 자원으로써 챗지피티의 텍스트 생성 능력 평가 -

루카 펠레그리니

독어독문학과

성신여자대학교 대학원

이 연구는 이해 가능한 입력을 통해 **A2** 수준의 독일어 학습자의 독일어 습득을 돕기 위해서 **ChatGPT**를 활용한 텍스트를 생성의 타당성 여부를 조사한다. 논문은 5장으로 구성된다. 2장은 이론적 틀에 관한 것으로, 입력 가설과 **AI**의 교육적 활용 방안에 대해 다룬다. 이 장에서는 언어 습득과 학습의 차이를 분석하고, 입력 가설을 깊이 탐구하면서, 이해 가능한 입력에 있어 무엇보다도 읽기가 중요함을 강조한다. 또한 **ChatGPT**가 기술적 측면과 기능 및 교육 콘텐츠 생성에 있어서 갖는 잠재력에 주목하여, 특히 교육적 맥락에서 **AI**를 활용하는데 있어 그 가능성과 한계를 다룬다. 3장은 실증적 연구를 위한 단계로서, **ChatGPT**를 통해 생성한 텍스트의 적절성을 평가하기 위한 기준으로 텍스트 길이, 가독성, 어휘의 적절성을 제시한다. 4장에서는 **ChatGPT**가 이러한 기준에 따라 독일어 **A2** 수준의 텍스트를 적절하게 생성해낼 수 있는지를 평가한다. 마지막으로 5장에서는 연구 결과를 종합하면서 **AI**를 통해 생성한 텍스트가 외국어 학습에 있어 갖는 의의를 평가한다. 이를 통해 **ChatGPT**와 같은 **AI**의 교육적 활용이 외국어 습득에 얼마나 기여할 수 있는지에 대한 통찰을 제공하고 있다.

키워드: **ChatGPT**, 독일어 습득, 이해 가능한 입력, 입력 가설, 텍스트 생성