Università degli Studi di Padova

Dipartimento di Studi Linguistici e Letterari

Corso di Laurea Triennale Interclasse in
Lingue, Letterature e Mediazione Culturale (LTLLM)
Classe LT-11

Tesina di Laurea

Data-driven learning and artificial intelligence: a comparison between two approaches to the creation of language teaching materials

Relatore
Prof. Katherine Ackerley

Laureando
Stefano Mallqui
n° matr.2045154 / LTLLM

Anno Accademico 2023 / 2024

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis advisor Katherine Ackerley for her invaluable patience and feedback. I am also extremely grateful to my Erasmus tutor, Alice Equestri, who generously supported me when I was studying abroad. Many thanks to my classmates, old friends and new friends I made along the way for their moral support and for all the unforgettable experiences I was able to made thanks to them. I would like to extend my sincere thanks to the international friends I have made in Leicester for giving me the best period of my life and for helping me grow personally. I would be remiss in not mentioning my family, especially my parents for always caring, helping and motivating me in doing something that I like.

TABLE OF CONTENTS:

Introduction

In the following academic thesis, I will delve into the complex and articulated system of learning English as a second language, analysing different approaches and methodologies for future language teachers. The purpose of this research is to discover which methodological approach to language acquisition is going to be more effective and reliable over a long period, whether it is with the implementation and usage of artificial intelligence technologies or with data-driven learning tools.

Through careful analysis, I will compare a human-made exercise made with DDL tools with an exercise generated with artificial intelligence for a hypothetical ESL class of students. The results will eventually be taken into consideration and re-examined furthermore in the future because artificial intelligence is developing and improving day by day, and it is hard to make projections of how good it is going to be in a couple of years.

The first chapter is going to explain in-depth the concept of data-driven learning and why it is relevant in language acquisition. I will break down the meaning of the term, how it works and how it is possible to use data to enhance and improve in learning a language. Furthermore, I am also going to analyse the advantages as well as the disadvantages for the learner in different contexts. This first chapter is going to be essential in understanding the arguments and the topics that I will cover in other chapters.

In the second chapter, I will dive into the world of artificial intelligence, also called AI, and the potential use of it with data-driven learning. To be more specific, I will start by defining and explaining what AI is and how it could be used in teaching environments. I will start by defining and explaining what AI is and how it could be used in teaching environments, especially when it comes to teaching languages. Furthermore, I will discuss the impact of AI on DDL methodologies and what are the future prospects for AI by analysing articles and research on the matter. As I said previously, it is going to be hard to determine how AI is going to develop and the impact it will have on us in the future.

This final chapter aims to conduct a comparative analysis between data-driven learning exercises and exercises generated by ChatGPT to determine which of these is more effective in language learning. Since I can not test these exercises with real students, I decided to establish a real-world scenario with a class of 13-year-old ESL students with an A2 level of English. In this chapter, I will discuss the similarities and differences that I have found in designing these exercises, also dwelling on what future researchers should consider if interested in delving deeper into the topic.

1. Data-driven learning

This first chapter aims to provide a brief but complete explanation of data-driven learning. I will examine its principles, applications, and potential obstacles, revealing how data-driven insights are revolutionizing education. Moreover, I am also going to investigate the potential of data-driven learning to create tailored and responsive learning environments suited for the learner's needs through an analysis of previous research.

1.1 Definition

In the 1990s, Tim Johns was the first researcher to introduce the data-driven learning approach, which marked a significant turning point in the field of education (Johns 1986). This innovative approach was the first one that allowed learners to access and analyse vast amounts of data to drive their learning experience. As Johns (1997:101) stated, by using this educational system, every student should turn into Sherlock Holmes and start investigating the language rules by themselves. The term data-driven learning (DDL), as written in Liang et al. (2023:81) "is a method of learning a foreign language based on corpus data which provides new ideas for the reform of foreign language teaching methods.". With DDL, Zhang (2022:2) affirms that "Learners are independent builders of knowledge, and the learning process is a knowledge construction process of self-creation and self-motivation.". To be more specific, Hadley and Charles (1094:131) declare that data-driven learning "is a student-centered inductive method of language learning, in which learners explore grammar and vocabulary issues using a corpus.". Essentially, when using a DDL methodology, the students undertake the role of researchers, with a specific focus on observing language usage based on corpus data, whereas teachers play the vital role of guiding students and providing them with the necessary resources and training to carry out their research effectively. By leveraging the available resources and applying their analytical skills, students can extract meaningful insights from the corpus data and gain a deeper understanding of language usage patterns (Johns 1991).

The term corpus (plural: corpora) refers to a curated assembly of texts gathered for specific purposes. Unlike a random collection of texts, a corpus is systematically organized based on predetermined criteria (Cheng 2011:3). According to Zhang (2022:3) "Corpus is a large-scale collection of written and spoken natural language materials that serve as a warehouse for storing language materials, also known as a language database.". It is important to note that corpus linguistics is a systematic study of language usage based on real-life examples, rather than relying on intuition or guesswork. This can include language samples from a range of sources, such as newspapers, magazines, spoken conversations, and more (Costas Gabrielatos 2005:2). A concordance, a computer program that serves to search for and analyse a specific word or phrase, represents an indispensable tool for accessing a corpus. The Keyword in Context (KWIC) format constitutes the most prevalent method for representing concordance lines. This format presents the target word or phrase in the centre of each line, with the surrounding context stored alphabetically to the right and left. The combination of a corpus and concordance provides learners with abundant examples to deduce the meanings and patterns of a target word or phrase (Asmaa Al-Mahbashi & Noorizah Mohd Noor:15). Moreover, as mentioned by Conza-Armijos & Celi-Celi (2023:3), "When retrieving results with KWIC [...] learners may find the structural analysis more manageable provided that the interface is visually appealing.". Furthermore, throughout the KWIC concordance format, learners are able to 1) identify keywords relevant to their inquiry and conduct related searches; 2) classify the results and organize concordance lines by sorting them to the left or right to identify patterns of usage; 3) derive their own hypotheses based on the results (Johns 1991: 4).

Nowadays, a multitude of electronic corpora are readily available for research and educational purposes, rendering the laborious and time-intensive task of manually compiling texts and analyses obsolete. The increasing popularity of corpus research has led to a growing interest in academic resources and the incorporation of corpora in educational methods (Rasikawati 2020:6). As stated by Luo (2016:3) "the size of a corpus may determine the effectiveness of DDL activities.". With that being said, I believe it is crucial to discuss the corpus that is most commonly cited. According to

Boulton and Vyatkina (1990:15-16), the BNC (British National Coprus) is the most widely used corpus appearing in 24% of papers, surpassing COCA (Corpus of Contemporary American English) which is cited in 20%. COCA has seen a significant increase in usage, being cited in nearly one-third of all studies between 2016 and 2019, amounting to 32% (Davies 2009:159-187). Despite their evident appeal, the potential of corpora and tools such as WordAndPhrase or SkELL, which are software to analyse language usage patterns, remains largely untapped. Only 19 studies utilize the web directly as a "corpus", maintaining a consistent 5% usage rate throughout the entire period. Furthermore, emerging trends indicate the emergence of new, larger, and specialized corpora, although there appears to be a lack of substantial or innovative types of corpus.

1.2 Importance and applications

In explaining the importance and significance of DDL, it is crucial to reveal the results of this methodology by looking at the discoveries of other researchers. Before doing that, I would like to draw attention to 5 key points about data-driven learning that make this approach extraordinary.

1. According to linguistic theory, language is complex, dynamic, and interactive, making it challenging for learners to understand its rules. Corpus linguistics has provided many insights into language patterns, which can help learners deal with authentic language use. DDL helps learners recognize these patterns, making it easier for them to learn the language (Tomasello 2005; Taylor 2012).
2. Learning theory suggests that learners find it easier to identify patterns rather than rules. The human brain is programmed to detect patterns, making it easier for learners to identify and learn patterns in language use. DDL promotes this skill, helping learners transfer it to new contexts and promoting learner autonomy (Aston 1998:13).
3. Psycholinguistic theory highlights the importance of pattern induction, which reduces cognitive load and makes it easier for learners to process meaning. DDL provides authentic language input and helps learners identify patterns, making it easier for them to notice important details. Chunking is also important in

language learning, and DDL helps highlight this skill (Sweller, Ayres & Kalyuga 2011; SCHMIDT 1990).

4. Second language acquisition (SLA) research has recommended a shift from a focus on meaning to including bottom-up processes such as form-focus. DDL offers a way to teach language that includes these processes, moving away from vocabulary lists and grammar exercises (Doughty & Williams 1998).
5. Finally, DDL can build on existing learner practices, such as using Google as a concordance for the Web as a corpus. Properly designed DDL activities can refine these practices and use them as a way to teach corpus work (Chinnery 2008; Kilgarriff & Grefenstette 2003).

Having said this, I can begin to reveal what has been discovered concerning the effectiveness of DDL. The recent research conducted by Ackerley (2017:13-14) on improving phraseology with DDL, has revealed positive effects both in the written production, as in the acquisition of a wider vocabulary of the majority of students that were taking part of the study. Another study made by Oktavianti (2015:55-63) on teaching vocabulary using the British National Corpus, has brought beneficial outcomes. Even on questionnaires, students from multiple studies said to have appreciated this new approach and found it easy and useful to work with authentic data by their own (Corino & Onesti 2019:11). According to Indra Nugraha, Miftakh and Wachyudi (2017:4) students that have experienced the implementation of DDL in class, have produced important testimonies: "This class was different from traditional English lesson; I can see the example of grammar use contextually; This type of learning is active rather than passive; I can see many more example sentences than in a dictionary; I had more interest in learning grammar through the task.".

I will discuss more about the advantages and disadvantages of DDL later in this chapter. Now, I would like to highlight the applications of this methodology with students of different levels and ages. According to multiple studies, it has been found that data-driven learning results in a substantial improvement among advanced and intermediate-level language learners (Boulton & Cobb 2017; Gordani 2013; Braun 2007; Sun & Wang 2003). However, according to other research, the impact of data-driven learning on intermediate-low-level students has brought mixed results (Boulton 2009a; Hadley

2002; Mizumoto and Chujo 2015; St & John 2001). In a study conducted by Boulton (2009b:50-51), it has been examined whether lower-level learners could effectively use authentic corpus data as a reference source without prior training. The results showed that using corpus samples led to better outcomes compared to traditional pedagogical resources, such as bilingual dictionaries and grammar manuals, that the learners were already familiar with. Both corpus data and pedagogical resources were found to be equally useful for recall purposes. However, the learners found authentic contexts presented in the form of multiple KWIC (Key Word In Context) concordances more manageable than longer contexts comprising one or more full sentences. These findings suggest that data-driven learning (DDL) could be beneficial for a wider range of learners than previously believed. DDL can be helpful for advanced learners with corpus training, but it still offers advantages even for learners at lower proficiency levels. Informal feedback from participants indicated that the primary challenge was not with the DDL approach or KWIC presentation, but rather with grappling with the complexity of authentic language usage. In addition, Hartle (2023:16) noted that introducing language tools, such as SkeLL, to learners early on, including postgraduate students embarking on their academic writing journeys, allows ample time for vocabulary development before potentially entering professional writing careers later in life. Such interfaces are readily accessible and user-friendly for learners, offering valuable language models in the generated output that can be leveraged to enhance personalized production. These interfaces can be viewed as lifelong learning resources.

1.3 Techniques and Methodologies

Having addressed the definition and explanation of the importance of data-driven learning, I can now discuss the techniques and methodologies that can be used in the classroom. Before revealing those, I would like to point out two main things that I think are usually taken for granted. First, as mentioned by Granger and Gilquin (2010) there is a wide range of DDL activities that can be created and used. Moreover, when designing the activities, the only limit the teacher has is his/her imagination. The second thing I wanted to highlight, is that, according to Boulton (2010), Thurstun and Candlin's Exploring Academic English (1997) has been an incredible resource of DDL activities ready to use. Lastly, before uncover how activities are structured, it is important to keep in mind that, as stated by Hirata and Thompson (2022:5) "before students can begin to analyse concordance lines effectively, appropriate learner training is needed". This is very essential, so much so that Avetisyan (2022) argues that teachers play a vital role in guiding students to effectively utilize language tools. However, before they can effectively assist students, teachers must first acquire the necessary knowledge and training. This preparation equips them to explain the functions and utilization of corpora in language learning with clarity and precision. By investing time in training for both teachers and students on using corpora and concordances, schools can maximize the potential of these tools to help students learn languages better, making learning a more engaging and effective experience for everyone involved (Smart 2014:1-3; Leńko-Szymańska 2017:1-4; Oktavianti 2015:55-57).

Since DDL activities are created with a concordance output, there are three main types of exercises that are possible: frequency lists, keyword-in-context (KWIC) lines and samples of texts (Papaioannou 2018). DDL activities are divided and categorised as 'hands-on' and 'hands-off' (Boulton 2017:15-36) or as 'hard' and 'soft'(Gabrielatos 2005b:1-37) or as a direct or indirect approach. The main difference is that with a direct data-driven learning approach, students can consult, access and analyse online corpus with the help of computers and special programs, whereas with a hands-off approach, students will not have the ability to navigate through a corpus by themselves, but they will receive an activity design by the teacher on handouts (Corino & Onesti 2019:3-4). The choice between using an approach instead of another, typically falls within the opportunity or not to access to a computer, or it could also be dictated by the language level of the student (Granger and Gilquin 2010:6-7). According to Boulton (2017) the difference between direct and indirect DDL is not simply based on the way it is delivered, but goes beyond that. The benefits of employing a hands-on DDL activity compared to a hands-off DDL activity, as I have previously hinted, are that students will acquire more flexibility, autonomy, lifelong learning, and long-term recall. A critical drawback could be that a lot of valuable classroom time can be wasted because of the lack of technical back-up or inappropriate searches and both teachers and learners are overwhelmed by simultaneously integrating fresh content (the corpora), advanced technology (the software), and innovative methodology (DDL) (Boulton 2017). When designing an activity with a soft DDL approach, teachers can perform searches beforehand, carefully curate and organize small data sets, and present them through tailored activities that match the learners' needs and abilities or use materials already prepared. The handouts usually have conventional activity formats, such as gap-fills, matching etc., offering a tangible and achievable goal in the short term. As a result, it can prove to be more inspiring than hands-on activities, and also more suitable for certain learning preferences (Boulton 2012).

According to Hadley (2002:15), after conducting tests, it was found that hard DDL methodologies were not suitable for low-level learners. While data-driven learning can be beneficial for learners at different learning stages, it has been found that it can be more effective for intermediate to advanced learners who have a foundational understanding of the language. For those who are new to it, DDL can be quite challenging because of the intricacies of corpus data, so beginners may benefit more from structured instruction that is tailored to their level (Boulton 2017). This finding aligns with one of the earliest instances exploring hands-off DDL implementation in the classroom, as outlined by Johns (1991). Johns developed a handout containing concordances featuring sample sentences illustrating various meanings of the modal "should". Following the examination of these concordances, students were tasked with identifying appropriate labels for the categories and potentially uncovering the relationships between them. From the advantages' perspective, this methodology offers several benefits. It can be applied immediately to all language-level students, it requires minimal or no corpus training, and students do not have to scroll down countless concordance lines focusing on the task provided by the teacher (Johns 1991:1-16). However, as I previously anticipated, the disadvantage of this type of approach is that students will not gain the knowledge to consult corpora, therefore they will be deprived of using them by themselves (Corino & Onesti 2019:3-4). An intriguing alternative approach, proposed by Charles (2007:11-12), involves students working on a computer during class and receiving a hard copy of the concordances as a record of their progress. The best part is that this hard copy can also be utilized for further study at home. I genuinely think this combination is really useful and effective in teaching languages with data-driven learning because students will utilize both hands-on and hands-off materials.

1.4 Benefits and Challenges

According to Boulton (2017:10-11), data-driven learning is slowly transforming education systems thanks to his innovative approach. Its primary advantage lies in its ability to provide unparalleled insights into language usage patterns and structures. By analysing corpora, teachers can develop a broad understanding of real-world language use for later structuring tailor exercises for the unique needs of their students. Since learners are being exposed to authentic language materials, DDL allow them to drastically develop their language proficiency. Furthermore, they also acquire the competence and language skills to allow them to be able to keep up a conversation in multiple real-world scenarios. Additionally, thanks to this exposure, it has been said that there is an improvement in the student's language knowledge, confidence and fluency (Chen 2011). A second advantage of using a data-driven learning methodology is that it promotes the cultivation of critical thinking and analytical skills among learners. By engaging with corpora analysis, learners develop the ability to evaluate and interpret language data, enabling them to make informed linguistic judgments and draw evidence-based conclusions (Azzaro 2012; Chang & Sun 2009; Tribble 2014). Moreover, as seen in O'Sullivan (2007) and Yoon (2014) studies, this set of skills extends beyond language learning, empowering students to apply critical thinking skills in various academic disciplines and different occasion. Lastly, data-driven learning serve an important role in exploring the variation and change of a language. Educators and students can investigate linguistic phenomena across different genres, registers, and contexts, gaining insights into sociocultural factors that influence language use. According to multiple researches (Boulton 2009b; Granger and Gilquin 2010; Lin & Lee 2015; Yoon 2011), this exploration help language learners to acquire a wider cultural awareness and understanding of the world around them.

Unfortunately, throughout various studies including those conducted by Mukherjee (2004), Lüdeling and Kytö (2008), and Leńko-Szymańska (2017), it was found that

DDL is facing some difficulties in being integrated into school, one of them being the fact that many educators lack of technological knowledge. While the availability of equipment in schools could be considered as a valid concern, there is a more complex issue at play: the lack of specific training for teachers. Not only the majority of them are unfamiliar with corpus linguistics and other related ICT (Information and Communication Technology), but they also lack the necessary skills to incorporate available resources into their teaching (Schaeffer-Lacroix 2019). To conclude, I would like to agree with what Corino (2009:5-7) stated saying that teachers should receive comprehensive training and experience in these areas so that they can effectively utilize these technologies to make lessons more engaging and to improve the learning outcomes of students.

1.5 Towards the deepening of artificial intelligence

Having now understood how important and useful data-driven learning is, I can now gradually explain what is artificial intelligence and how it can be integrated in education. Additionally, I will uncover both the advantages and challenges associated with employing artificial intelligence, as well as its potential to enhance DDL. From this careful deepening, I will later be able to compare and analyse data-driven learning exercises with exercises produced by artificial intelligence.

2. Artificial Intelligence (AI)

The following chapter aims to provide a brief but complete explanation of what artificial intelligence is, and its usability in language learning contexts. I will examine the impact that AI has on data-driven learning, highlighting the benefits and challenges of this possible integration. Moreover, I am also going to discuss ethical considerations and future trends and developments in artificial intelligence, as this is a very controversial issue.

2.1 Overview

The born of the term "artificial intelligence" dates back to 1945 during an academic conference held at Dartmouth College by John McCarthy, the founding father of artificial intelligence (Anyoha 2017). Artificial intelligence can be described as "The theory and development of computer systems able to perform tasks normally requiring human intelligence, such as, visual perception, speech recognition, learning, decision-making, and natural language processing." (2017:1) or as stated by Wartman and Combs (2018:1107), the term expresses the ability of computers and machines to replicate a virtual human mind that performs different tasks depending on the work, as if they were performed by a human being. According to Cardona, Rodríguez and Ishmael (2023:12), the term "human-like" not only describes modern computer technologies but also highlights their significant development in comparison to early edtech applications. Additionally, according to Timms (2016), artificial intelligence in the future will not only be present in computers used daily, but it will also take on different shapes and functions as it becomes integrated into our lives.

The way in which artificial intelligence works is particularly complex. According to Kiger (1970), it is a combination of large amounts of data with intelligent algorithms that allows artificial intelligence software to learn from patterns and features of the data. According to his research, and other recent studies, like the one conducted by CSU

Global (2021), it is mentioned that in order to emulate a human brain, AI software uses a combination of these different subfields:

1. Machine Learning: which aims to automate the development of analytical models to discover hidden information within the data without specific programming.
2. Artificial Neural Networks: which has the task of mimicking the interconnected structure of brain neurons, transmitting information between units to identify correlations and extract meaningful insights.
3. Deep Learning: that aims to use extensive neural networks and significant computational power to discover complex models within data, particularly in areas such as image and speech recognition.
4. Cognitive Computing: which plays the role of emulating natural, human interactions, including interpretation and response to speech.
5. Computer Vision: which uses pattern recognition and deep learning techniques to understand the content of images and videos, allowing real-time interpretation by machines.
6. Natural Language Processing: which involves the analysis and understanding of human language, facilitating appropriate responses and interactions.

As seen in Borana (2016) and in Marr (2023) the applications of artificial intelligence are many and various. It can be implemented in healthcare areas to facilitate doctors in improving patients care and outcomes throughout personalised treatment plans and easier medical image analysis, in financial sectors to help traders and buyers, in transportation industries to make travel routes better and to develop self-driving cars, in manufacturing areas to automate repetitive processes to build a product, in entertainment industries in order to help content creators and to make experiences more personal, in agriculture sectors with the aim of helping farmers grow crops in a smarter and more sustainable way, and in education systems where it can become a useful resource for teachers and students.

2.2 Artificial intelligence in education

According to Boulay (2020), AI can take three main roles in class. It can be used to assisting individual students, the whole class or the whole cohorts of students and as stated by Ahmad et al. (2023), the outcomes achieved through the utilization of artificial intelligence in class vary depending on its application. Focusing on foreign language teaching, artificial intelligence and technology plays a vital role in offering students new possibilities for language learning. Liu (2023:1-3), listed some applications of artificial intelligent for foreign language learning classes, and among them, there are speech recognition technology, machine translation technology, natural language processing technology and chatbot technology. Now I want to see these technologies more in detail as they all are relevant and important also in DDL.

Speech recognition technology, for example, has been proven to be a valuable tool to significantly improve students' communication skills. This technology allows language learners to receive immediate feedback on their pronunciation and intonation, which is crucial in developing their oral expression skills. As a result, students can become more confident in their ability to communicate in a foreign language and are more likely to engage in conversations with native speakers. Additionally, speech recognition technology can help learners improve their listening and comprehension skills. Students can use speech recognition software to listen to recordings of native speakers, and the software can identify any mistakes in their pronunciation. This helps students to hear the correct pronunciation and intonation of words and phrases, thereby improving their understanding of the language. Speech recognition technology also provides learners with the opportunity to practice their oral expression skills using training materials and oral practice. Students can record themselves speaking and compare their recordings to the original recordings of native speakers. This helps students to identify areas that need improvement and work on them to enhance their oral expression skills.

Aside from speech recognition technology, machine translation tools has also proved to help students better understanding foreign language texts and elevate their language proficiency. These tools can also optimize the accuracy and speed of translations, resulting in substantial time and effort savings. By incorporating these tools into their language learning process, students can effectively improve their language abilities and communication skills.

In Liu's research (2023:1-3) is also mentioned the natural language processing (NLP) technology that focuses on the interaction between computers and human language. Through its analysis of a student's performance across various language domains, it generates personalized learning materials. By analysing authentic language data, NLP facilitates a natural and effective learning process for students. With its potential to revolutionize foreign language teaching and learning, there is anticipation of more innovative applications as the technology continues to evolve. These advancements will undoubtedly assist students in acquiring foreign language skills with greater efficiency (Yang 2023, Yamamoto 2023).

On the other hand, chatbots, which are computer programs designed to simulate conversations with human users, have been proven to be a valuable resource for language learners seeking to enhance their skills. With interactive features, students can practice speaking naturally and receive instant feedback on their grammar, pronunciation, and vocabulary usage. Real-time translations, definitions, and conversational phrases are also readily available, making chatbots an ideal tool for those without access to native speakers or language tutors (Haristiani 2019:1-7).

With the implementation of AI in the educational field, it is increasingly prevalent to perceive that the role of teachers, schools, and leaders in education will change (Gocen & Aydemir 2020:13). According to Manyika et al. (2017:3-4) good teachers will continue guiding their students towards enhanced creativity, communication, and emotional intelligence, and unexpectedly, technology such as artificial intelligence and automation will foster greater humanity among individuals. Furthermore, as briefly stated by Haseski (2019), the use of artificial intelligence will make learning more personalized, provide effective experiences, enable students to discover their talents, improve their creativity, and reduce teachers' workload. On the contrary, the delegation of teaching responsibilities to computers is regarded as a peril in research on artificial intelligence (Mozelius & Humble 2019). Pedró (2019:12-15) emphasizes the importance of a dual-teacher model that utilizes artificial intelligence for individualized education. Teachers often spend a considerable amount of time on routine tasks, such as repeating information and answering frequently asked questions on various topics. With the help of digital AI-supported assistants in the classroom, teachers will have more time to focus on personalized student guidance and one-on-one communication. This will ultimately improve the overall quality of education by allowing teachers to provide more individualized attention to each student.

According to Liu (2023:3), when it comes to foreign language learning classes, artificial intelligence tools can be a useful resource as they bring exciting opportunities for students to access a wealth of language materials and exercises. Students can effectively improve their foreign language proficiency with the help of speech recognition and machine translation capabilities (Liu 2023:3). Moreover, it is also mentioned that one of the main advantages of AI technology lies in its ability to tailor educational content and teaching methods to suit the unique learning style and situation of every student, leading to personalized education and enhanced learning efficiency. By providing prompt and precise language materials and exercises, AI technology has the potential to assist students in acquiring foreign language proficiency. Gocen and Aydemir (2020:16-17) conducted an examination of the implications of AI for the future of education. They found that the capacity of AI to evaluate individuals' skills, identify specific educational needs of learners, and accommodate different learning styles has led to the development of personalized education. Additionally, Gocen and Aydemir (2020:16-17) also stated that artificial intelligence has the ability to reduce the administrative tasks of teachers and improve the overall workflow by facilitating faster decision-making and helping teachers plan better educational activities, thus raising standards and outcomes.

On the other hand, the disadvantages brought up throughout Liu's (2023: 3) research are related to the fact that even though artificial intelligence technology has the potential to enhance student's learning, it can not fully replace human teachers when it comes to certain key areas such as language expression and interpersonal communication. This is due to the fact that AI lacks the humanistic approach that is essential in making the learning experience more engaging. Moreover, it is also important to be aware of technical limitations and security concerns, such as inaccuracies in speech recognition and machine translation, and chatbot technology compromising personal information. Other drawbacks pointed out in the research of Gocen and Aydemir (2020:16), may involve promoting a rigid mindset towards cognition, diminishing the importance of humanistic values, categorizing individuals solely based on quantitative measures, favouring factual knowledge at the expense of well-rounded growth, removing the need for human intervention, raising concerns over data privacy and security, and weakening interpersonal connections.

2.3 Impact on Data-driven Learning

A recent research by Crosthwaite and Baisa (2023:2) revealed that data-driven learning is in danger of being outclassed and overshadowed by artificial intelligence technology, which essentially does the same thing. The only difference between them is that AI is doing it in a way that has finally captured the imagination of the public. For over two decades, corpora have been used to enhance the teaching and learning of languages. This was achieved indirectly, through the creation of dictionaries, wordlists or integration of corpus data into teaching materials, or directly, through learners' hands-on consultation of corpora. A fascinating fact that caught my attention from Crosthwaite and Baisa (2023:2) research is the assertion that is made.

> "Does this mean the end of DDL as we know it? Upon reflection, it certainly does mean the end of DDL as we know it if corpus linguists and DDL practitioners do not take steps to rectify ongoing issues that have continually plagued the field and begin to consider how GenAI may help us overcome them."

Before discussing the ways in which data-driven learning can be improved with the implementation of artificial intelligence technology, I am going to highlight some of the main advantages of DDL listed in Crosthwaite and Baisa (2023) research.

1. Knowing the data: One of the main advantages of using corpora as a research and teaching tool is the comprehensive understanding of the domain of texts from which the corpus data is derived. This is not the case with current large language models that power applications like ChatGPT. With corpora, we know the exact texts that are included in large general corpora such as the BNC2014

and the BAWE (Crosthwaite, Sanhueza & Schweinberger 2021:9-12). It is also possible to extract the complete texts from these corpora if needed. This knowledge is important for research and teaching purposes as it allows learners to take complete ownership over the corpus used for their queries. CorpusMate, for instance, has a “citation” function that provides details about the corpus from which the text was derived, including the title of the text and a link to the corpus. The ability to use DIY corpora, where learners can select the texts themselves, is a considerable advantage in this regard (Charles 2012:94).

2. Authenticity: It is important to note that there are some differences between language data produced by humans and the language produced by GenAI, which uses a statistical procedure to generate sentences. Although GenAI can create grammatically correct sentences, they may not always be appropriate for the given context or register, and may not be commonly used in actual writing or conversation. It may be necessary to adjust the prompts to reduce these issues when using GenAI. For second language learners who cannot easily verify whether a given output is accurate for the target language, it is generally more reliable to use authentic corpus data - that is, language data created by humans (Crosthwaite & Baisa 2023:2).
3. Replicability: GenAI applications use complex statistical procedures to generate text. However, end-users currently cannot see the statistical procedures that lead to the generated text, nor replicate them. Even if one could replicate the procedures, the answers are randomly sampled, leading to a unique answer for each subsequent identical query. On the other hand, corpora allow for easy replication of a given finding with the same query on the same data, producing "hard evidence" that word X belongs with word Y, for example. This evidence is incredibly powerful for language learners and their teachers, especially if the finding can be replicated in different corpora. Such replication can be done time and time again without limitation (Crosthwaite & Baisa 2023:2).
4. Multimodality: Several recent corpus tools offer multiple ways of accessing corpus data, such as colored concordances, statistical tables (like collocation scores), and visual charts and maps of relationships between words and grammatical units (Sinclair & Rockwell 2016). These improved functionalities

of corpus tools specifically target and highlight patterns in corpus data, making it easier for users who may find traditional concordancing difficult to use. Some pedagogical corpora, such as SACODEYL (Pérez-Paredes & Alcaraz-Calero 2009:56-73), use video and audio files along with concordancers. Currently, most GenAI tools can generate tables or detailed images from text prompts. However, it can be challenging to implement this within a chatbot context, and it often requires integrating one tool with another, which can be difficult for non-technical users.

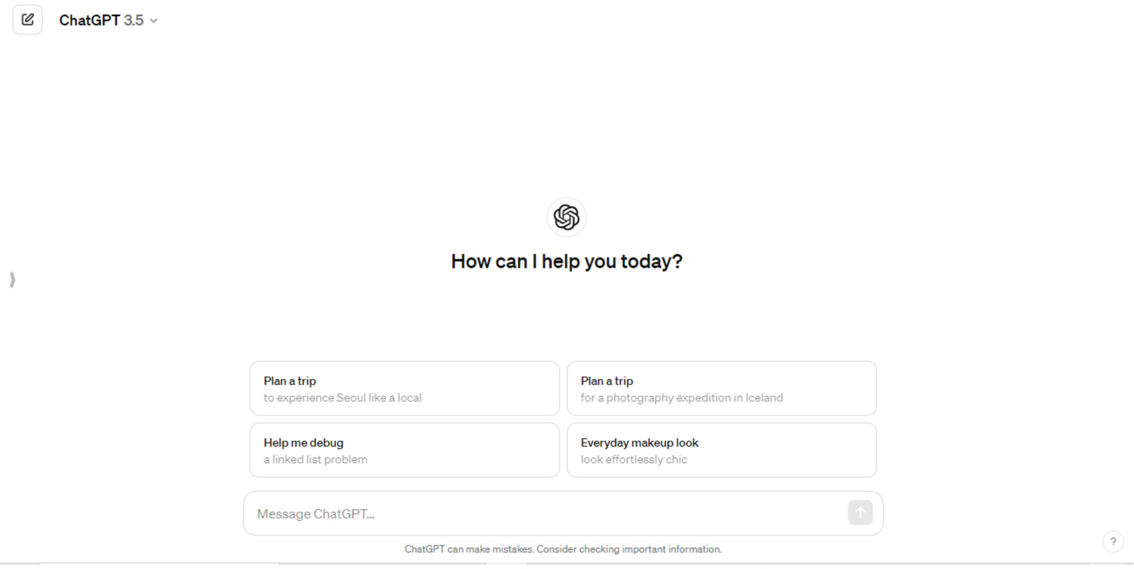
5. Safety: When it comes to using GenAI tools and companies at primary and secondary education levels, there is often a concern about the lack of transparency regarding the usage of user data. Any data provided by users who are not yet in tertiary education is subject to strict ethical and legal safeguards, especially personal data, data on curricula, assessments, and more. Educational institutions are reluctant to allow even their staff access to ChatGPT due to these concerns, let alone younger students. However, most corpus tools require very little user data, such as only initial registration details. Therefore, corpus linguists who want to promote corpus use in schools can emphasize that corpus consultation is currently a "safe" choice (Crosthwaite & Baisa 2023:2).
6. Illusions: While ChatGPT is a helpful AI-powered assistant, it is important to note that there may be limitations in its output accuracy. For instance, its ability to generate non-Latin script languages may not be as strong as its ability to comprehend them (Bang et al. 2023:2). Additionally, there may be instances where ChatGPT generates terms that are not included in its training data (Shen et al. 2023). It is worth noting that unlike with corpus data, users have no control over the algorithm and are unable to access the data.
7. Active vs. passive learning: The literature on the “L” in DDL suggests that active learning, which involves constructing knowledge through exploration and practice, is required for successful corpus consultation. Concordancing, a task that involves identifying patterns in language data, requires significant inductive learning processes (Sun & Wang 2003:83-94). Crosthwaite and Baisa (2023:2) affirm that ChatGPT promote inductive learning by using patterns and associations learned by the model to generate text or provide relevant

> information when a user interacts with it. However, it is unclear to what extent ChatGPT or the learner is responsible for the induction process. This can result in a significant risk to the learning process of students as some users may simply copy and paste ChatGPT output without actually learning anything.

While data-driven learning through corpora analysis offers unparalleled insights into linguistic patterns and structures, Crosthwaite & Baisa (2023:3) suggest that integrating AI into the equation introduces a dynamic element that enriches the learning process even further. More specifically, Crosthwaite & Baisa (2023:3-4) research delineates seven ways in which data-driven learning can be improved thanks to artificial intelligence.

As I mentioned in the first chapter, using DDL can be challenging and requires both teachers and students to prepare before they can navigate the corpora on their own successfully (Schaeffer-Lacroix 2019). With the implementation of artificial intelligence it can become easier so much so that the technical knowledge required to consult large language data is reduced. It is no longer necessary to use complex syntax to isolate parts of speech in our corpus, as now, it is simple to request what is needed..

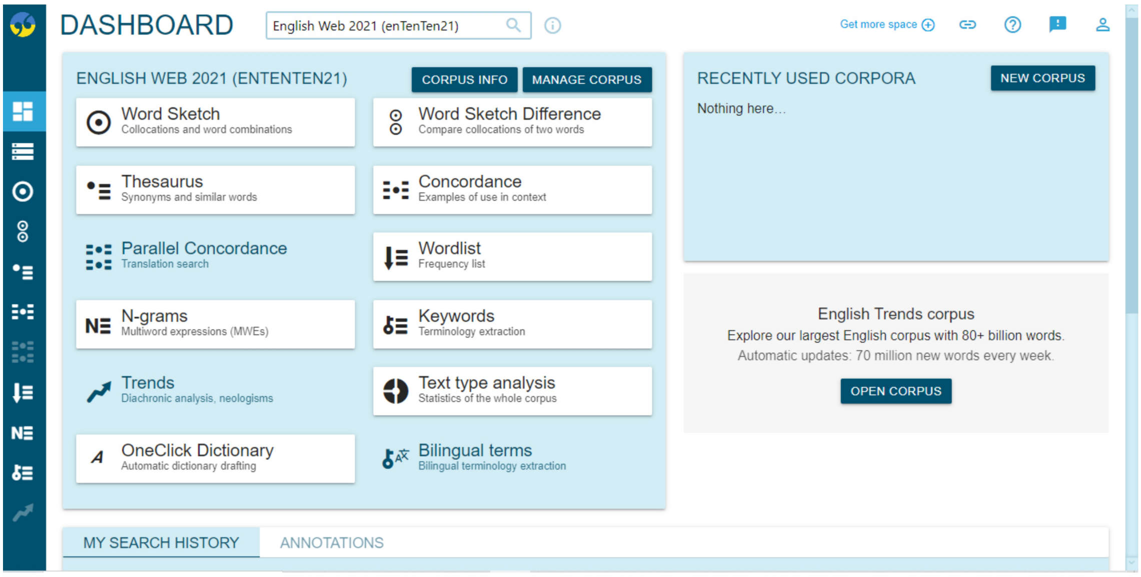
Besides DDL training sessions, another most common complaint is related to the complexity of the corpus tools, particularly the user interface (UI). According to Crosthwaite and Baisa (2023:3) up to 2019 (31%) of corpus tools used highly complex UIs such as English-corpora.org. Although these corpus tools are primarily used for research rather than teaching, attempts to simplify their UIs have been largely unsuccessful, such as SketchEngine's recent "overhaul" (Boulton & Vyatkina 1990:80). In contrast, GenAI applications like ChatGPT are popular among the general population because they are incredibly user-friendly in both form and function. Here are two images that display both the graphical user interface of ChatGPT (3.5 version) in Figure 1, and the graphical user interface of Sketch Engine in Figure 2. One one side, there is simplicity; on the other, there is functionality.



How can I help you today?

ChatGPT can make mistakes. Consider checking important information.

Figure 1: ChatGPT graphical user interface (OpenAi 2022)



Explore our largest English corpus with 80+ billion words. Automatic updates: 70 million new words every week.

Figure 2: Sketch Engine graphical user interface

Another problem of DDL can be given by the fact that its rigidity can represent a difficulty for those who do not have the necessary proficiency to interpret the concordance results accurately. However, GenAI applications have the capacity to accurately differentiate language that is meant for advanced users of the target language and simplify it by using more straightforward structures and vocabulary. This feature has been identified as a game-changer by teachers of English as an additional language/dialect in Australia, especially for classes with mixed levels of proficiency. The ability of GenAI to generate results from almost any domain, register, or language cannot be overstated and can expand the scope of DDL beyond its current focus on tertiary academic English language.

Another important aspect to take into account from Crosthwaite & Baisa (2023:3) research, is the sheer size of the latest language models, like ChatGPT (which uses GPT-4), are trained on billions or even trillions of tokens. The ability to efficiently query models of this magnitude at high speed online was previously unheard of, even with the most advanced corpus tools available.

Furthermore, another advantages of most of AI technologies, is that they are able to refine future outputs by using previous inputs to the model. These "chats" are saved for future use, making it easier to track how learners engage with corpora. While screen recordings and query logs have been used in the past to capture longitudinal data, it is not currently possible to easily capture this information using existing corpus tools. (Kotamjani et al. 2017:61; Pérez-Paredes et al. 2011:236-238) According to Crosthwaite & Baisa (2023:3), researchers are likely to revisit user interactions with these language models and use chatlogs as evidence of "learning". This will be a major research methodology as researchers investigate the extent to which GenAI use promotes language acquisition and better learning in general, something that DDL research has yet to achieve.

Additionally, throughout various studies such as Ma et al. (2022:6-14), it appeared that it requires a teacher with in-depth knowledge of language, content, and pedagogy to transform corpus findings into effective teaching materials for Data-Driven Learning (DDL) lesson planning. However, the ability of GenAI/ChatGPT to effortlessly convert language-focused queries into a teaching task, assessment item, or even a complete lesson plan is truly remarkable. For instance, it can convert a query like "what are some nouns that fill slot X in this sentence" into a list of nouns that can be used to create multiple-choice questions or a lesson plan for 2nd graders to acquire these nouns. Crosthwaite & Baisa (2023:3) consider this technology as a game-changer for mainstream pre-tertiary teachers who have limited time and resources.

The seventh and final point that is addressed is the ability to afford a financed working group always updated. Here, Crosthwaite & Baisa (2023:3-4) are referring to the fact that individuals or small teams at universities often build language tools with limited funding or as a hobby, resulting in outdated or unfinished tools. Companies like OpenAI frequently update their software, and while we can not compete with their speed and scale, collaborating with them can be advantageous.

Before discussing the ethical considerations of artificial intelligence, I would like to highlight a recent research conducted by Lin (2023) regarding ChatGPT and DDL because it not only pointed out how important it is to formulate great prompts in order to get high-quality results but also tested whether the latest ChatGPT version was able to generate the same type of information as DDL tools. To be more specific, Lin tried to replicate concordance lines and frequency lists, but unfortunately, it turned out that, despite many attempts, ChatGPT failed to generate the desired results.

2.4 Ethical Considerations

According to research conducted by Gocen and Aydemir (2020:19), AI has the potential to simplify and streamline processes in education. However, it is not a cure-all solution. When using AI tools, it is essential to consider legal issues such as responsibility for AI tool actions, ownership of AI-generated products, and privacy and security concerns. Laws must be created that keep pace with advances in technology and address specific AI tool issues in education. According to Cardona, Rodríguez & Ishmael (2023:54) when developing AI, it is crucial to be careful and think about its potential effects, and most importantly, to prevent any harm, some legal guidelines should be established for using AI in education. It may seem obvious, but it is important to assess how well certain tools serve educational priorities. Sometimes, people get carried away with the potential of technology and adopt a "let us see what the tech can do" attitude. This can lead to a weakened focus on goals and cause us to choose models that do not fit our priorities well. Furthermore, according to Cardona, Rodríguez and Ishmael (2023:56), the integration of AI in the education field requires collaboration among various stakeholders, such as educational leaders, teachers, students, and families. To ensure the effectiveness and adaptability of AI models in educational settings, it is crucial for developers and regulatory authorities to prioritize transparency and accountability. Moreover, it has been reiterated how effective government oversight can play a vital role in guaranteeing that these models are functioning optimally and meeting the changing needs of the education sector. The regulations laid out in the Blueprint for an AI Bill of Rights (2022:5-7) state that you, as a user, should be protected from unsafe or ineffective systems. Algorithms should not discriminate against you, and systems should be designed and used in a fair and equitable way. You should also be protected from abusive data practices with built-in protections, and you should have control over how data about you is used. Additionally, you should be informed when an automated system is being used, and understand how and why it contributes to outcomes that affect you. You should also have the option to opt out when appropriate, and have access to a person who can quickly address any issues you may encounter. However, being able to provide all of this can be challenging and requiring a lot of time and laws, as shown on the examples within the Blueprint for an AI Bill of Rights (2022:29-52). To conclude, Cardona, Rodríguez & Ishmael (2023:60), also emphasize the importance of reviewing and considering the regulations related to student and family data privacy laws such as FERPA, CIPA, and COPPA. This is particularly important in light of the increasing use of new and emerging technologies in schools, including AI-enabled learning technologies. As new situations arise, it may also be necessary to take into account laws such as IDEA, which is related to individuals with disabilities education. To conclude, it is important that in the future, specific ethical rules are guaranteed to ensure that language students can benefit from DDL with the implementation of artificial intelligence.

2.5 Future Trends and Developments

According to the study conducted by Liu (2023:4-5), the future of foreign language teaching is set to experience significant and innovative advancements with the development of artificial intelligence. As AI technology continues to rapidly developing, it is anticipated that its integration into foreign language education will become increasingly intelligent, interactive, personalized, and efficient. In Liu's

(2023:4-5) research, it is also stated that AI technology has the potential to improve learning effectiveness and produce better results in foreign language learning by analysing data on individual students' learning situations, strengths, and weaknesses. By simulating real-life scenarios with artificial intelligence tools, teachers can help students practice their foreign language skills and apply the grammatical rules learned in class, making language learning more relevant, engaging and effective. Moreover, foreign language learning classes with the implementation of artificial intelligence will benefit from the possibility of using virtual reality, and game-based learning materials to make classes even more interactive and engaging. In addition, it is also mentioned that AI language learning tools such as voice recognition technology, personalized feedback, and robotic assistance empower students to seamlessly attain fluency in foreign languages. Based on Liu's research, it is clear that AI will play an increasingly important role in foreign language education in the coming years, but on the other hand, this development will bring both opportunities and challenges that will require careful consideration and practical application. As AI technology continues to evolve and improve, foreign language learning will become more effective, engaging, and accessible to a wider range of learners.

2.6 Towards an analytical comparison

Having now understood how artificial intelligence can actually be a useful resource in the future also for educational purposes, I can now move on to analyse and compare data-driven learning exercises with exercises produced by ChatGPT for a hypothetical class of ESL (English as a second language) learners. By comparing these approaches, I can discover nuanced differences in learning efficiency, adaptability, and the potential impact on student learning.

3. Comparison of Exercises

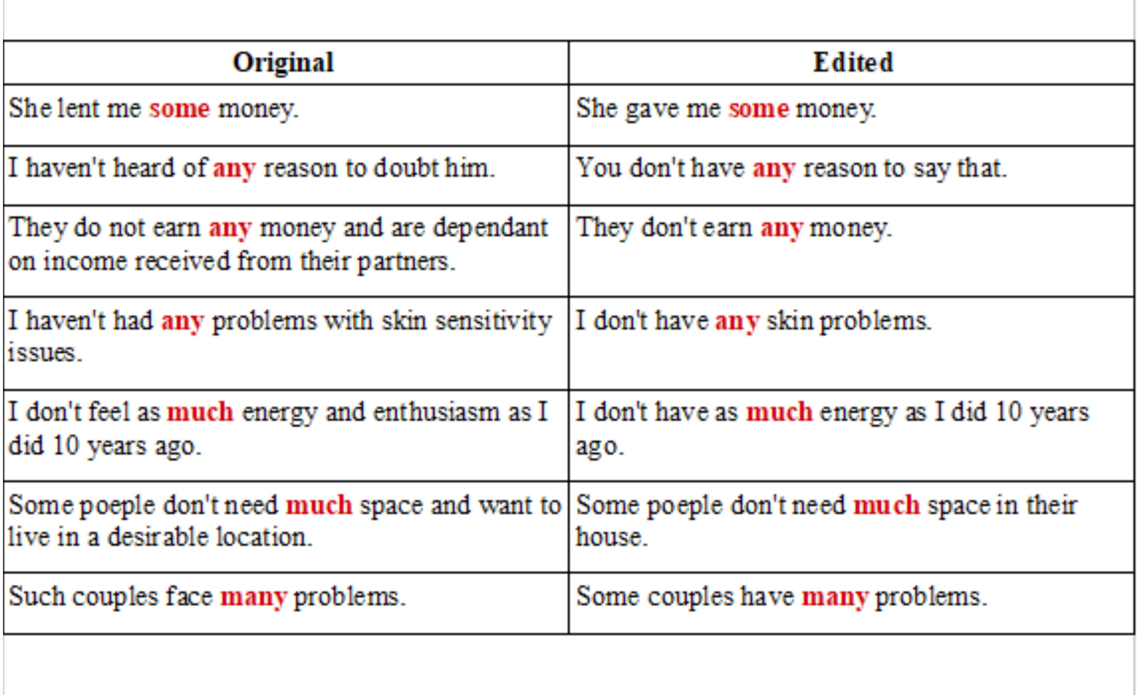
This final chapter aims to conduct a comparative analysis between data-driven learning exercises and exercises generated by ChatGPT to determine which of these is more effective in language learning. Apart from this, it will also focus on uncovering the respective differences and strengths of each method. The exercises will be structured for a hypothetical English as a second language class for 13-year-old students with an A2 level of English. The activities will focus on countable and uncountable nouns, a topic they have already covered. At the end of the chapter, I will summarise all the collected data and the results obtained from this comparison.

3.1 DDL Method

The DDL activities I created using Sketch Engine focus on countable and uncountable nouns. I am presenting these two exercises with an indirect data-driven learning approach because I do not expect 13-year-old students to have the necessary knowledge to navigate a corpus by themselves. For these tasks, I have decided to use Sketch Engine to search sentences featuring "some", "any", "much", and "many" using the English Web Corpus 2021 (enTenTen21). First, I went to the "Concordance" section to start searching for the terms I needed. Instead of displaying KWIC, I selected to see the entire sentences with the selected term. Next, I used the frequency lists to find phrases where these terms were used with common and simple words suitable for students at an A2 English level. Finally, I selected some sentences to create two sets of concordance lines for the exercises. Unfortunately, despite my attempt to use only sentences provided by Sketch Engine, some phrases that I have selected need some adjustment to allow young ESL learners to understand and being able to complete the tasks individually. The data-driven learning activities presented consist of a true/false task and another exercise in which students need to fill out a table and complete some statements with the correct option.

3.2 Modifying concordance lines

As I previously said, some sentences that I have extrapolated from Sketch Engine needed some adjustment to be more suitable for ESL students with an A2 level of English. As you can see in Figure 3, I am displaying the original sentences and how I edited them.



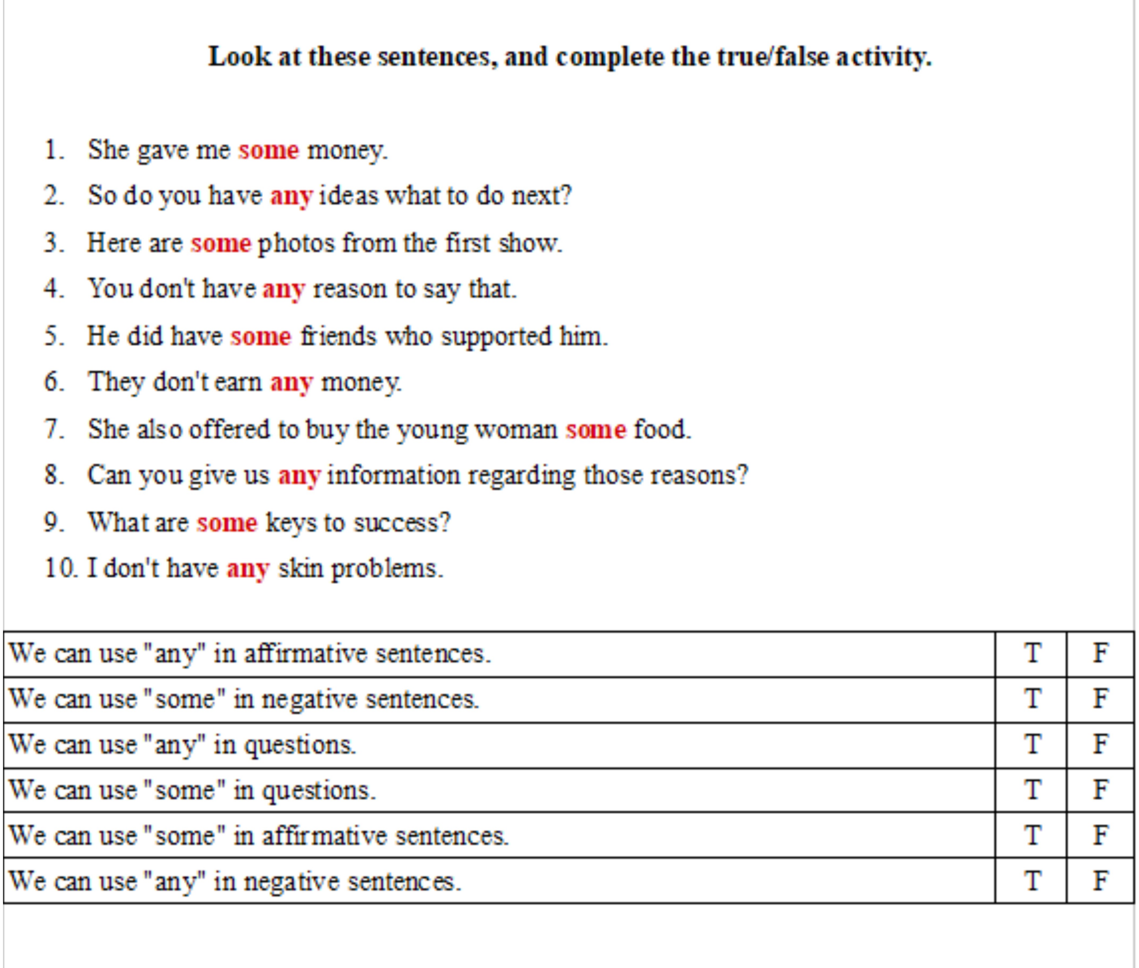
Original	Edited
She lent me **some** money.	She gave me **some** money.
I haven't heard of **any** reason to doubt him.	You don't have **any** reason to say that.
They do not earn **any** money and are dependant on income received from their partners.	They don't earn **any** money.
I haven't had **any** problems with skin sensitivity issues.	I don't have **any** skin problems.
I don't feel as **much** energy and enthusiasm as I did 10 years ago.	I don't have as **much** energy as I did 10 years ago.
Some poeple don't need **much** space and want to live in a desirable location.	Some poeple don't need **much** space in their house.
Such couples face **many** problems.	Some couples have **many** problems.

Figure 3: Adjusting concordance lines

In order to make these sentences suitable for a class of 13-year-old ESL students with an A2 level of English, I decided to change some complex vocabulary from the original phrases with words that are more common and easy to understand. Moreover, on many occasions, I decided to shorten the sentences to make the students focus more on the terms "some", "any", "much" and "many". Lastly, when needed, like in "Such couples face many problems." apart from using "have" instead of "face", I also gave context to the sentence by changing "Such" with "Some". Having now discussed the method and the changes that I have adopted, I can start presenting the two DDL exercises.

3.3 Task 1

The first set of tasks will focus on the use of “some” and “any”. The DDL exercise I created is designed to help students discover when it is appropriate to use "some" and "any" by presenting example sentences and having them complete a true/false activity. The sentences listed in the exercise are sourced from Sketch Engine, and some of them have been edited to be more suitable for students with an A2 level of English. The exercise is as follows:



Look at these sentences, and complete the true/false activity.

1. She gave me **some** money.
2. So do you have **any** ideas what to do next?
3. Here are **some** photos from the first show.
4. You don't have **any** reason to say that.
5. He did have **some** friends who supported him.
6. They don't earn **any** money.
7. She also offered to buy the young woman **some** food.
8. Can you give us **any** information regarding those reasons?
9. What are **some** keys to success?
10. I don't have **any** skin problems.

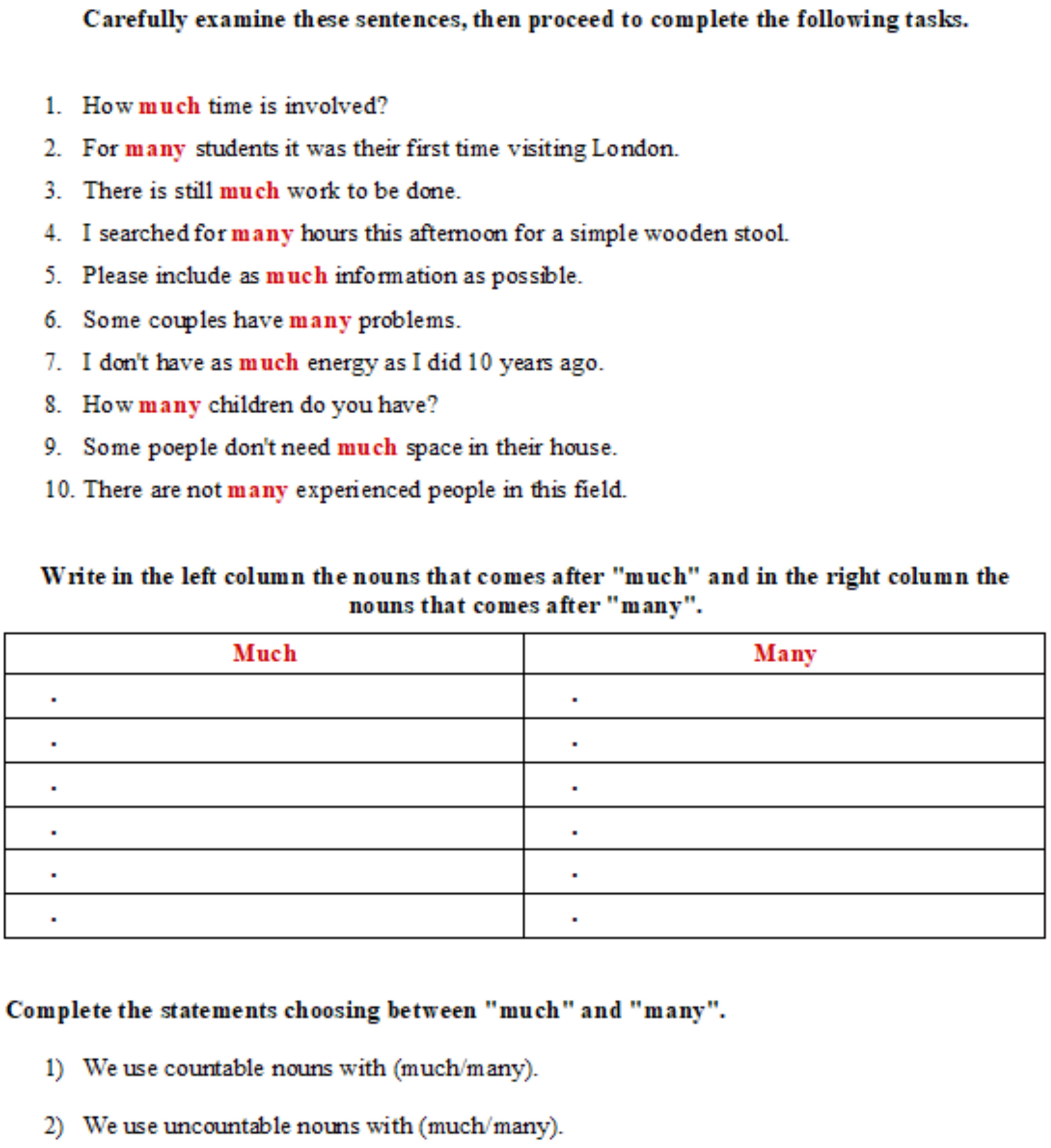
We can use "any" in affirmative sentences.	T	F
We can use "some" in negative sentences.	T	F
We can use "any" in questions.	T	F
We can use "some" in questions.	T	F
We can use "some" in affirmative sentences.	T	F
We can use "any" in negative sentences.	T	F

Figure 4: DDL Exercise on "some" and "any".

To complete this exercise, students should carefully examine all sentences and look for patterns or instances where "some" and "any" are used before attempting the true/false activity.

3.4 Task 2

The second DDL exercise I created is designed to help students discover when it is appropriate to use "much" and "many" with countable and uncountable nouns. The exercise is as follows:



Carefully examine these sentences, then proceed to complete the following tasks.

1. How **much** time is involved?
2. For **many** students it was their first time visiting London.
3. There is still **much** work to be done.
4. I searched for **many** hours this afternoon for a simple wooden stool.
5. Please include as **much** information as possible.
6. Some couples have **many** problems.
7. I don't have as **much** energy as I did 10 years ago.
8. How **many** children do you have?
9. Some poeple don't need **much** space in their house.
10. There are not **many** experienced people in this field.

Write in the left column the nouns that comes after "much" and in the right column the nouns that comes after "many".

Much	Many
.	.
.	.
.	.
.	.
.	.
.	.

Complete the statements choosing between "much" and "many".

1) We use countable nouns with (much/many).
2) We use uncountable nouns with (much/many).

Figure 5: DDL Exercise on "much" and "many"

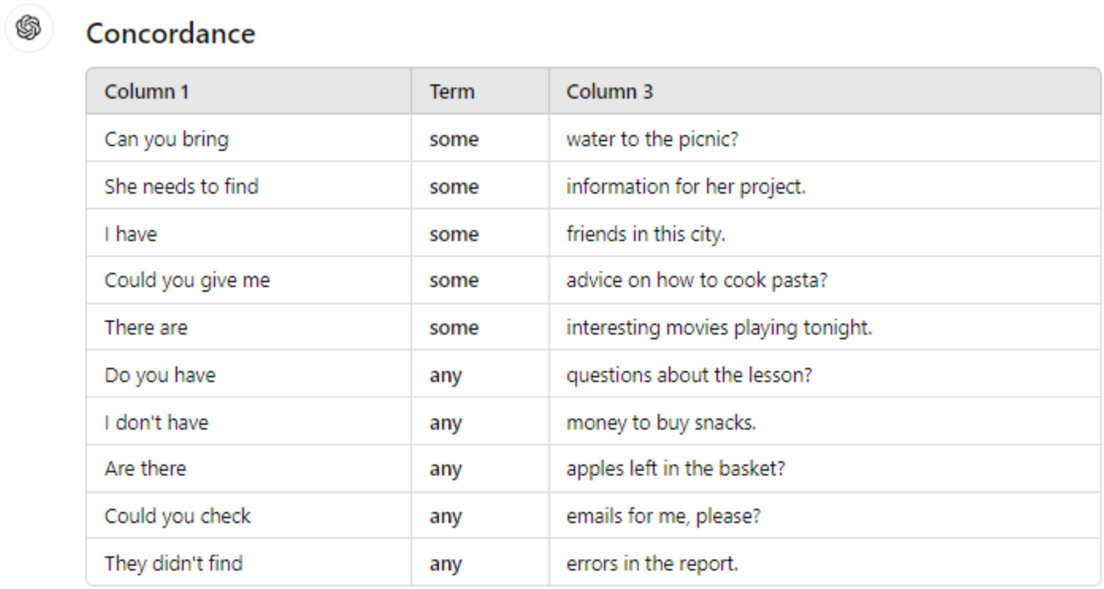
In order to complete this exercise, students should fill out a table with the nouns that come after "much" and "many". After doing this, students should understand by examining the table that "much" is used with uncountable nouns and "many" with countable nouns. This will allow them to complete two final statements.

3.5 AI Method

The exercises generated with artificial intelligence on countable and uncountable nouns are going to be developed by ChatGPT 3.5 version. To be more precise, I am going to ask ChatGPT to design activities similar to DDL exercises. To do that, I will give ChatGPT similar input and instructions as those present in Lin's (2023) study. I will conduct several tests with multiple inputs in order to get activities designed for the learner's level of English and also to see if AI can produce better tasks than a teacher, offering more suitable questions to guide the students in their inductive, discovery-based learning. I am going to ask ChatGPT to realise some concordance lines with the terms "some", "any", "much", and "many" and generate an inductive exercise suitable for 13-year-old ESL students with an A2 level of English. Before looking at the exercises generated, I want to reflect on the attempts that took me to get those results.

3.6 Failed attempt to generate an DDL exercise with AI

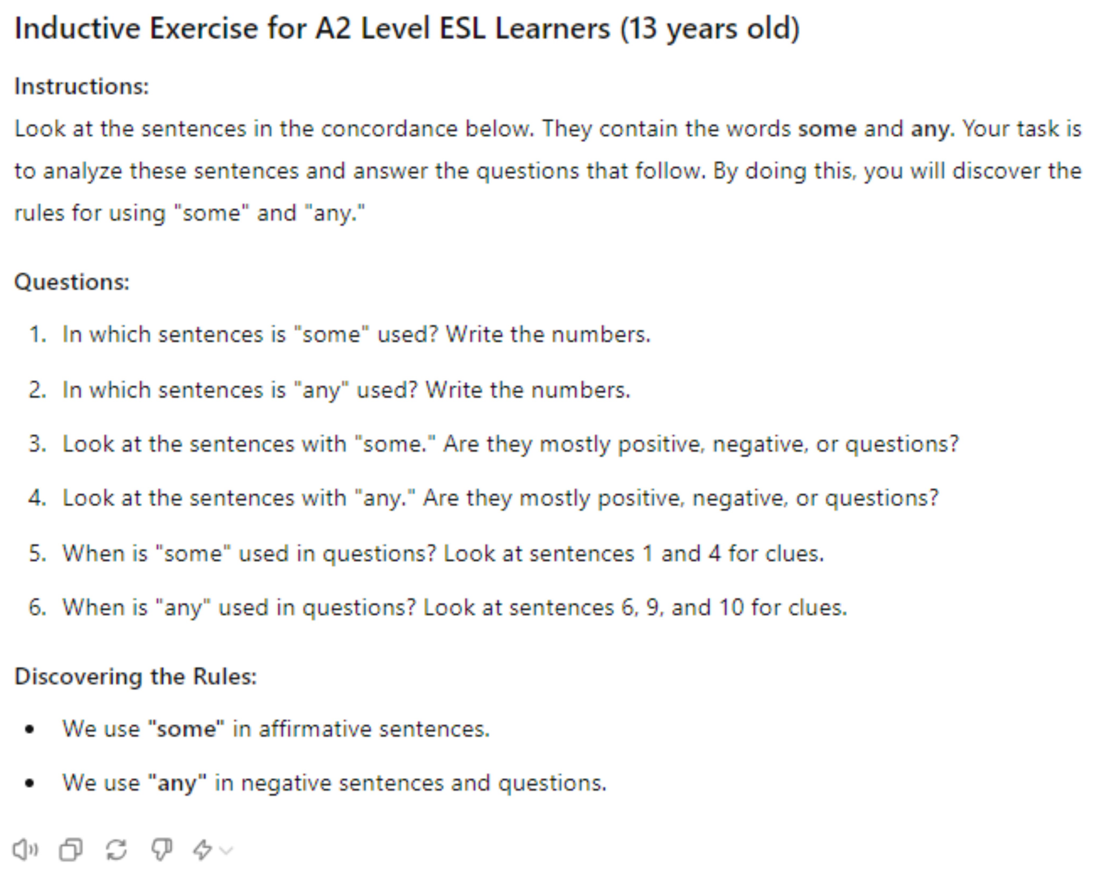
From the tests I did, it seemed to me that in order to get something specific and well done by ChatGPT, it is necessary to modify several times the indications you give it. Regarding the exercises with countable and uncountable nouns, it took me a bit of time to get the results I wanted. Now I'm going to show you an example of exercise that have not satisfied me for some reason. The input I gave to ChatGPT is the following: "Make a concordance with 10 sentences using "some" and "any." These sentences should show how the terms are used in English. Follow these steps when making the sentences: split each sentence into three parts and put them in three separate columns. Column 1 should have the words before the term, column 2 should have the term itself, and column 3 should have the words after the term. Make sure both column 1 and column 3 have at least 3 words unless the term is at the beginning of the sentence. The term must be bold. After this, design an inductive and easy exercise for 13-year-old ESL learners with an A2 level of English. The exercise should be similar to a data-driven learning exercise, with the student understanding when to use "some" and "any" looking through the concordance lines generated in the table. You do not need to display the answers to the exercise." As you can see, this input gives specific direction on how to build the concordance, but allows ChatGPT to design an activity based on some general indications. It is possible that this is why the concordance in Figure 6 is adequate, but the activity in Figure 7 is not convincing me.



Concordance

Column 1	Term	Column 3
Can you bring	**some**	water to the picnic?
She needs to find	**some**	information for her project.
I have	**some**	friends in this city.
Could you give me	**some**	advice on how to cook pasta?
There are	**some**	interesting movies playing tonight.
Do you have	**any**	questions about the lesson?
I don't have	**any**	money to buy snacks.
Are there	**any**	apples left in the basket?
Could you check	**any**	emails for me, please?
They didn't find	**any**	errors in the report.

Figure 6: First Attempt ChatGPT Exercise (Concordance) (OpenAi 2022)



Inductive Exercise for A2 Level ESL Learners (13 years old)

Instructions:
Look at the sentences in the concordance below. They contain the words **some** and **any**. Your task is to analyze these sentences and answer the questions that follow. By doing this, you will discover the rules for using "some" and "any."

Questions:

1. In which sentences is "some" used? Write the numbers.
2. In which sentences is "any" used? Write the numbers.
3. Look at the sentences with "some." Are they mostly positive, negative, or questions?
4. Look at the sentences with "any." Are they mostly positive, negative, or questions?
5. When is "some" used in questions? Look at sentences 1 and 4 for clues.
6. When is "any" used in questions? Look at sentences 6, 9, and 10 for clues.

Discovering the Rules:

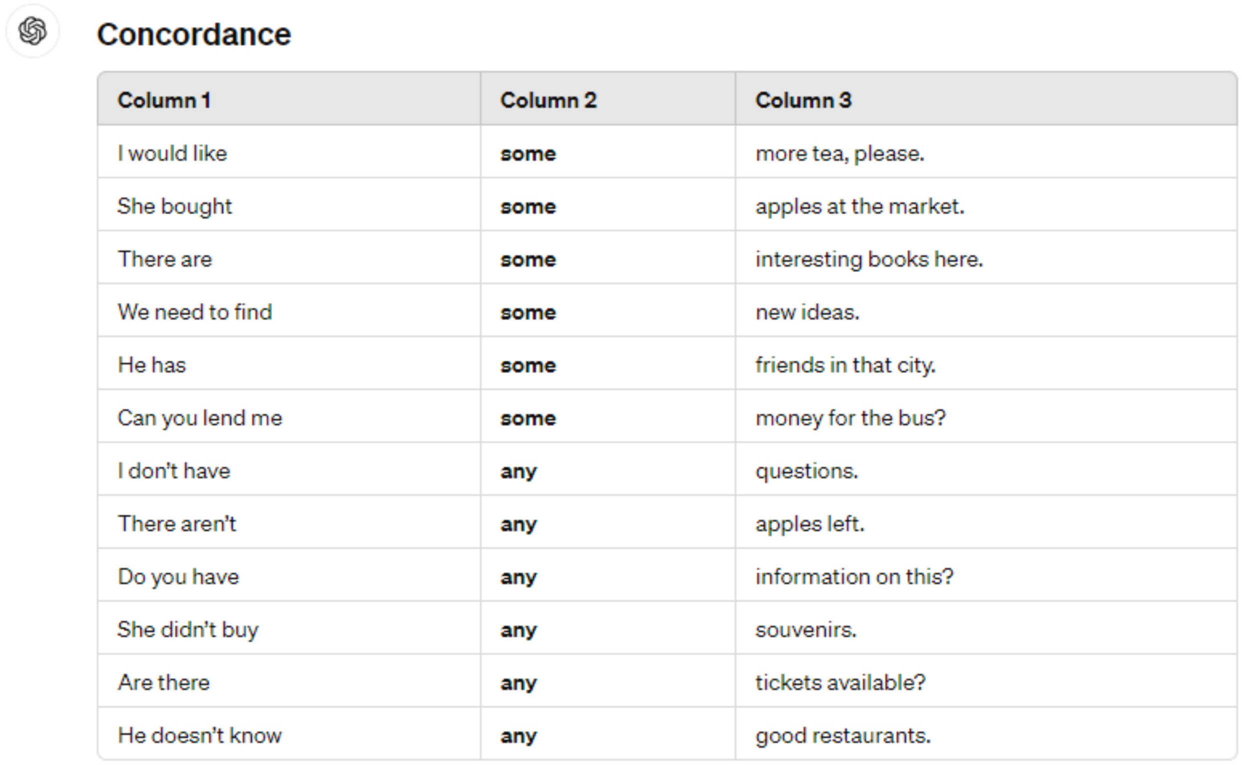
- We use **"some"** in affirmative sentences.
- We use **"any"** in negative sentences and questions.

Figure 7: First Attempt ChatGPT Exercise (Activity) (OpenAi 2022)

The exercise generated by ChatGPT displayed in Figure 7, which should imitate a DDL activity, is structured with just some questions that the students have to answer. Apart from the questions, this activity does not seem like an introductory exercise. By presenting all these questions, students may feel intimidated by the amount of commitment required, whereas, if the same activity was a multiple-choice exercise, I think it would look less like an exam and more like an introductory exercise. Now that I have shown you an example of ChatGPT's capabilities, I will present two activities that I will later compare with the DDL exercises.

3.7 Task 1

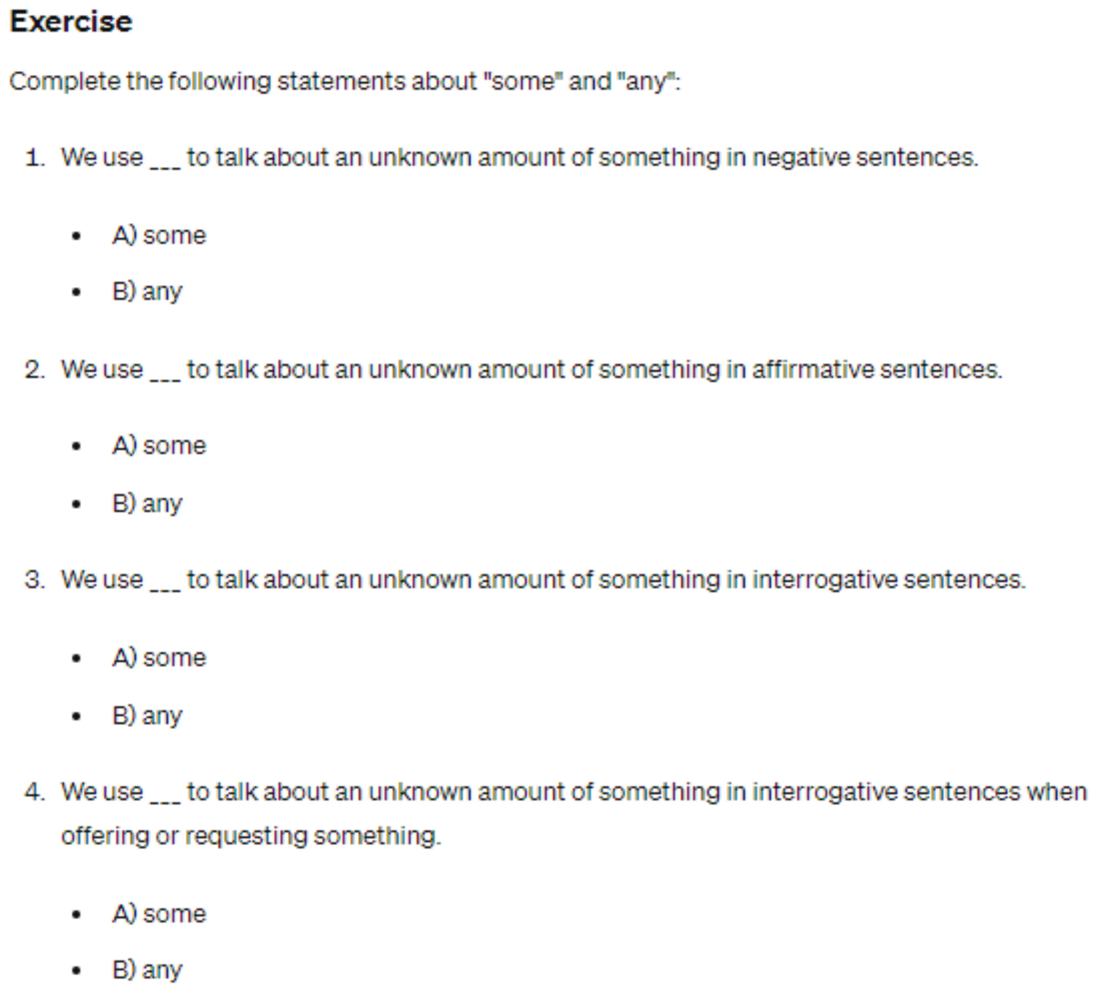
For the first exercise generated with ChatGPT, the input I typed is the following: "Generate a concordance consisting of 12 sentences with the terms "some" and "any". The sentences should illustrate how the terms are typically used in English. The sentences selected should be designed for students with an A2 level of English. In presenting these phrases, you should follow some directions. First of all, each sentence should be split into three parts and presented in 3 separate columns. Column 1 presents the words before the term, column 2 presents the term itself, and column 3 presents the words after the term. Besides that, column 1 and column 3 must both contain at least 3 words, except when the term is the first word of the sentence. On top of that, the term must be bold. Once you have done that, generate an exercise where students have to complete four statements about "some" and "any". For example: "We use... to talk about an unknown amount of something in negative sentences." under, option A is some and option B is any, "We use... to talk about an unknown amount of something in affirmative sentences." with option A some and option B any, "We use... to talk about an unknown amount of something in interrogative sentences." under option A is some and option B is any, "We use ... to talk about an unknown amount of something in interrogative sentences when offering or requesting something." under, option A is some and option B is any.". The exercise generated with ChatGPT appears as follows:



Concordance

Column 1	Column 2	Column 3
I would like	**some**	more tea, please.
She bought	**some**	apples at the market.
There are	**some**	interesting books here.
We need to find	**some**	new ideas.
He has	**some**	friends in that city.
Can you lend me	**some**	money for the bus?
I don't have	**any**	questions.
There aren't	**any**	apples left.
Do you have	**any**	information on this?
She didn't buy	**any**	souvenirs.
Are there	**any**	tickets available?
He doesn't know	**any**	good restaurants.

Figure 8: Concordance on "some" and "any" generated with AI (OpenAi 2022)



Exercise

Complete the following statements about "some" and "any":

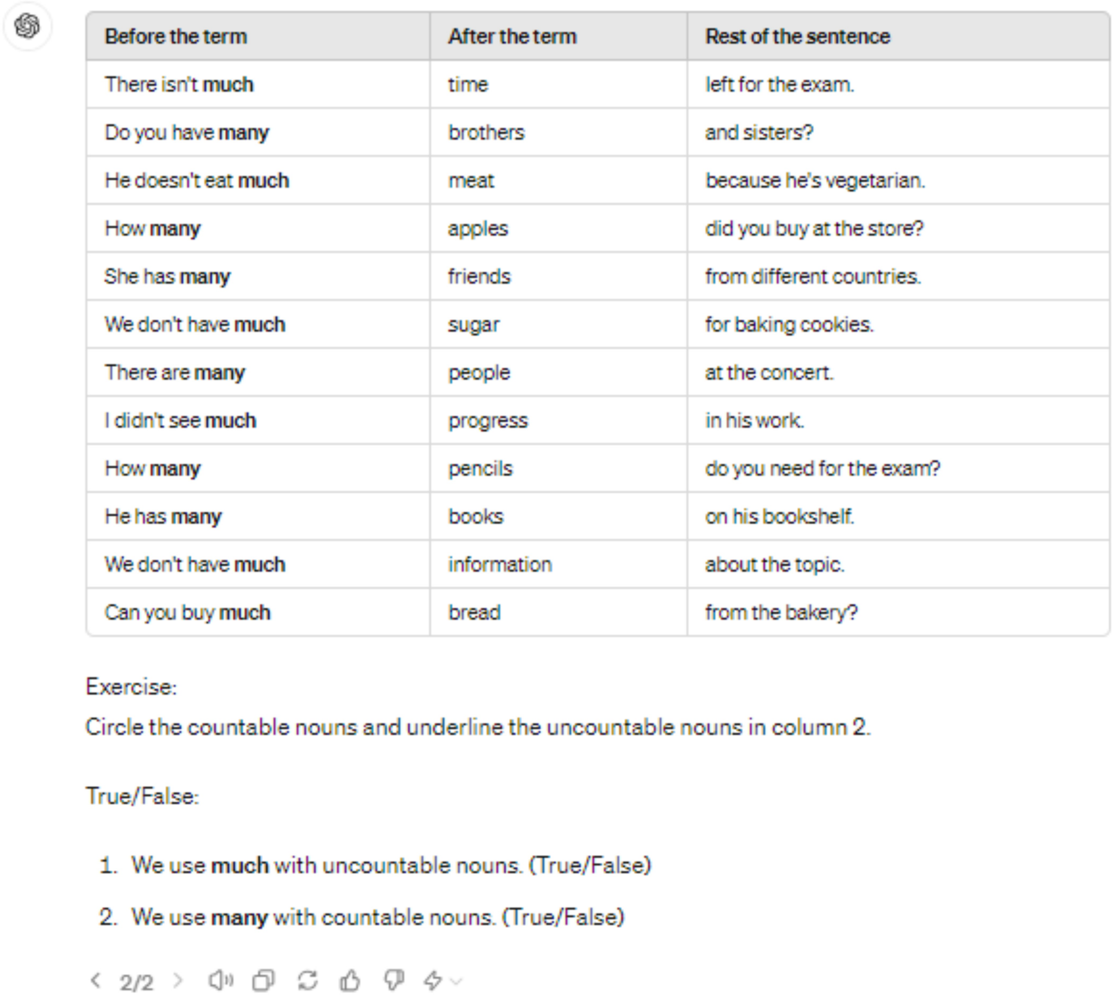
1. We use ___ to talk about an unknown amount of something in negative sentences.
   - A) some
   - B) any
2. We use ___ to talk about an unknown amount of something in affirmative sentences.
   - A) some
   - B) any
3. We use ___ to talk about an unknown amount of something in interrogative sentences.
   - A) some
   - B) any
4. We use ___ to talk about an unknown amount of something in interrogative sentences when offering or requesting something.
   - A) some
   - B) any

Figure 9: Exercise on "some" and "any" generated with AI (OpenAi 2022)

To complete this exercise, students should carefully examine all sentences and look for patterns or instances where "some" and "any" are used before attempting to complete the statements.

3.8 Task 2

For the second exercise generated with ChatGPT, the input I typed in is the following: "Generate a concordance consisting of 12 sentences with the terms "much" and "many". The sentences should illustrate how the terms are typically used in English. The sentences selected should be designed for students with an A2 level of English. In presenting these phrases, you should follow some directions. First of all, each sentence should be split into three parts and presented in 3 separate columns. Column 1 presents the words before the term and the term itself, column 2 presents the word after the term, and column 3 presents the rest of the sentence. Besides that, column 1 and column 3 must both contain at least 3 words, except when the term is the first word of the sentence. On top of that, the "much" and "many" must be bold. Later, create a short exercise in which students circle the countable nouns and underline the uncountable nouns in column 2. Afterwards, the students need to complete a true/false task. The statements for this activity could be "We use much with uncountable nouns." and "We use many with countable nouns.". The exercise generated with ChatGPT appears as follows:



Before the term	After the term	Rest of the sentence
There isn't **much**	time	left for the exam.
Do you have **many**	brothers	and sisters?
He doesn't eat **much**	meat	because he's vegetarian.
How **many**	apples	did you buy at the store?
She has **many**	friends	from different countries.
We don't have **much**	sugar	for baking cookies.
There are **many**	people	at the concert.
I didn't see **much**	progress	in his work.
How **many**	pencils	do you need for the exam?
He has **many**	books	on his bookshelf.
We don't have **much**	information	about the topic.
Can you buy **much**	bread	from the bakery?

Exercise:
Circle the countable nouns and underline the uncountable nouns in column 2.

True/False:

1. We use **much** with uncountable nouns. (True/False)
2. We use **many** with countable nouns. (True/False)

Figure 10: Exercise on "much" and "many" generated with AI (OpenAi 2022)

To complete this exercise, students should carefully determine whether the nouns that come after "much" and "many" are countable or uncountable and after doing that, they will be able to complete the true/false task.

3.9 Discussion

Starting from the similarities I found, it seems that ChatGPT successfully managed to replicate the inductive learning approach found in DDL tasks. As a matter of fact, in all the exercises, the students have to examine some sentences taken as examples and look carefully for patterns in how the terms are being used in different contexts. Another particular aspect of these exercises is that they are all with an adequate and appropriate vocabulary for 13-year-old ESL learners with an A2 level of English, but while ChatGPT has immediately provided simple sentences for the students, some concordance lines obtained by Sketch Engine had to be edited. This characteristic could be considered a negative aspect when it comes to producing a DDL exercise for low-intermediate-level of English learners, but it is important to bear in mind that also ChatGPT has its downsides.

Although I do not have a real class of ESL students to test these exercises, and therefore I cannot demonstrate their effectiveness and validity, I can discuss the main differences I encountered while creating these tasks. While the design process of the DDL exercises using Sketch Engine went smoothly with only a few adjustments, creating the exercise with AI was a completely different experience. Using indications similar to those in Lin's (2023) study, ChatGPT was able to generate sentences displayed in a well-organized concordance table with the terms in bold. The main issue I found with ChatGPT was realizing an activity similar to a DDL exercise, suitable for 13-year-old ESL students with an A2 level of English. In my opinion, Figure 7, where ChatGPT generates a task that attempts to replicate a DDL exercise, demonstrates that despite providing a general prompt like mine, the best that artificial intelligence has managed to produce so far was simple questions. I did not consider that exercise for my comparison because instead of being presented as an introductory exercise on "some" and "any", it looks more like an exam.

One last thing that is worth noticing is that despite the responsiveness of ChatGPT in producing multiple concordance lines with a specific term, it is crucial to remember that, on the other hand, a corpus like the one I used in this comparison, in this case, the English Web Corpus 2021 (enTenTen21), offers teachers the opportunity to choose sentences from a vast source of native and not native English resources. For the purpose of this study, I decided to address these exercises to an imaginary class of 13-year-old ESL students with an A2 level of English, but if using DDL for learners with a better knowledge of the language, with a B2 for example, the sentences inside the corpus will not require many adjustments as the students will comprehend most of them. The sentences generated by AI look similar to each other, but it is difficult to determine whether this is because of the equivalent input I gave to ChatGPT or because of specific algorithms.

3.10 Future perspectives

With just my research, I was not able to determine which between data-driven learning and artificial intelligence exercise is better suited for ESL students with a low intermediate level of English. Due to the rapid advancement of artificial intelligence and data-driven learning tools, it is crucial to conduct ongoing and up-to-date research on this subject to keep up with the changing landscape and potential new outcomes. I invite researchers to test both DDL and AI exercises with real students to obtain more information about their effectiveness, while also getting some feedback and opinion from the learner's point of view. To conclude, I think future researchers should also bear in mind the possibility that integrating data-driven learning and AI technology could eventually result to be the most effective way to teach English as a second language, as the strengths of DDL correspond to the weaknesses of AI and vice versa. These details offer hope for developing more responsive DDL tools and AI technologies capable of using vast amounts of language data and information.

4. Conclusions

Despite I was not able to determine which among DDL and AI exercises is more effective in language learning, this comparative research has allowed me to discover some interesting facts. To recap what I did, I first discussed in-depth about data-driven learning and artificial intelligence and how they are used in language teaching, and later I compared and examined two sets of exercises. Besides focusing on the similarities and differences, I also discussed the design process of these exercises, since it is completely different between them.

It is undoubtedly true that this study has provided valuable information about DDL exercises and those generated with AI, however, it is necessary to acknowledge its limitations. One big limitation of this research is the absence of real English as a second language students testing the tasks. If real learners had taken part in the study, in addition to the data collected in this research, I could have discovered and compared many more aspects. Furthermore, the lack of longitudinal data hinders the evaluation of the sustained effectiveness of teaching methodologies. Recognizing and proactively dealing with these methodological limitations can lead researchers to advance knowledge and improve methodologies, leading to a more clear comprehension of the phenomena being studied.

Future studies should test the exercises with two small classes of ESL students with an A2 level of English. Before the experiment, learners should take an exam to determine their language knowledge. Later, in one class, students should solve DDL activities, while in the other, AI technologies like ChatGPT should generate the tasks. After collecting all the exercises from the two classes, the data obtained should be analysed and compared. A final exam, similar to the one attempted before the experiment, should be taken from both classes to discover the class that has improved more. Moreover, after discovering this, it would be appropriate to interview students about their experiences and feelings of taking part in the experiment. Additionally, a few months after the interviews, in order to find out which methodological approach is better suited for

English learners, the two classes should be tested again. This last step is important to determine which of the two approaches is most targeted at lifelong learning.

Since this theme is still partially undiscovered, this thesis serves as a springboard to invite further researchers to deepen this intriguing theme. In this study, I found that both data-driven learning exercises and those generated with artificial intelligence have their respective strengths and limitations when it comes to improving English learning for students with an A2 level of English. Throughout this dissertation, I have not only managed to highlight the fact that both types of exercises can be further improved, but I have also paved the way for those interested in deepening and exploring what methodological approaches can facilitate students learning.

Bibliography

Ackerley, Katherine. 2017. Effects of corpus-based instruction on phraseology in learner English. *Language Learning & Technology ISSN* 21(3). 195–216. http://llt.msu.edu/issues/october2017/ackerley.pdf.

Ahmad, Kashif, Waleed Iqbal, Ammar El-Hassan, Junaid Qadir, Driss Benhaddou, Moussa Ayyash, Ala Al-Fuqaha, et al. 2023. *Data-driven artificial intelligence in education: A comprehensive review*.

Alejandro Curado Fuentes. 2017. Form-focused data-driven learning for grammar development in ESP contexts. *Revista de Lenguas para Fines Específicos* 23. 12–30. https://doi.org/10.20420/rlfe.2017.155.

Allan, Rachel, Terry Walker & Virginia Langum. 2023. Data-driven learning: Tools, approaches, and next steps. *Nordic Journal of English Studies* 22(1). 1–12.

Anyoha, Rockwell. 2017. The history of artificial intelligence. *Science in the News*. Harvard University. https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/.

Asi, Natalina, Akhmad Fauzan, Jean Seraf Yaspis, Stepanus Saputra, Ferry Lui & Ibnu Haikal Salasa. 2021. EFL STUDENTS' LONG-TERM PRACTICE OF DATA-DRIVEN LEARNING. *Exposure Journal* 390(2). 390–401. https://journal.unismuh.ac.id/index.php/expo.

Asmaa Al-Mahbashi & Noorizah Mohd Noor. The effect of data driven learning on receptive vocabulary knowledge of yemeni university learners. *The Southeast Asian Journal of English Language Studies* 3(3). 13–24.

Asti Ramadhani & Endah Lestari. 2022. *THE EFFECT OF DATA DRIVEN LEARNING TO IMPROVE VOCABULARY IN THE SECOND YEAR OF UNIVERSITY STUDENTS*.

Aston. 1998. Learning English with the BNC. *godzilla.sslmit.unibo.it*. https://godzilla.sslmit.unibo.it/~guy/barc.htm. (4 March, 2024).

Avetisyan, Marine. 2022. *LANGUAGE CORPORA AND DATA-DRIVEN LEARNING IN SECOND LANGUAGE ACQUISITION*.

Azzaro, Gabriele. 2012. Phrasal Verbs through DDL. *Ricerche di Pedagogia e Didattica -Journal of Theories and Research in Education* 7(2). (7 March, 2024).

Bang, Yejin, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Holy, Lovenia Ziwei, et al. 2023. *A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity*. (26 March, 2024).

Borana, Jatin. 2016. Applications of Artificial Intelligence & Associated Technologies. 64–67.

Boucher, Philip. 2020. *STUDY Panel for the Future of Science and Technology EPRS | European Parliamentary Research Service*. (24 March, 2024).

Boulay, Benedict . 2020. Artificial Intelligence in Education: where are we now? - Benedict du Boulay, University of Sussex. *www.youtube.com*. https://www.youtube.com/watch?v=54-yKBv6xeU. (25 March, 2024).

Boulton, Alex. 2009a. Data-driven learning: reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics* 35(1). 81–106.

Boulton, Alex. 2009b. Testing the limits of data-driven learning: Language proficiency and training. In *ReCALL*, vol. 21, 37–54. https://doi.org/10.1017/S0958344009000068.

Boulton, Alex. 2010. Data-driven learning: Taking the computer out of the equation. *Language Learning* 60(3). 534–572. https://doi.org/10.1111/j.1467-9922.2010.00566.x.

Boulton, Alex. 2012. Hands-on / hands-off: Alternative approaches to data-driven learning. https://hal.archives-ouvertes.fr/hal-00503034.

Boulton, Alex. 2017. Data-driven learning and language pedagogy. In *Language and Technology*, 1–12. Springer International Publishing. https://doi.org/10.1007/978-3-319-02328-1_15-1.

Boulton, Alex & Tom Cobb. 2017. Corpus use in language learning: A meta-analysis. *Language Learning*. Blackwell Publishing Ltd 67(2). 348–393. https://doi.org/10.1111/lang.12224.

Boulton, Alex & Nina Vyatkina. 1990. Thirty years of data-driven learning: Taking stock and charting new directions over time. *Language Learning & Technology* 25(3). 66–89. http://hdl.handle.net/10125/73450.

Braun, Sabine. 2007. Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL* 19(3). 307–328. https://doi.org/10.1017/s0958344007000535.

Bush, Vannevar. 1945. *A SClENllST OF THE FUTURE RECORDS EXPERIMENTS WITH A llNY CAMERA FlllED WITH UNIVERSAL-FOCUS LENS. THE SMALL SOUARE IN THE EYEGLASS AT THE LEFl SIGHTS THE OBl ECT AS WE MAY THINK A TOP U. S. SCIENTIST FORESEES A POSSIBLE FUTURE WORLD I N WHICH MAN-MADE MACHINES WILL START TO THINK*. (24 March, 2024).

Cardona, Miguel A., Roberto J. Rodríguez & Kristina Ishmael. 2023. *Artificial Intelligence and the Future of Teaching and Learning Insights and Recommendations*. (24 March, 2024).

Chang, Wen-Li & Yu-Chih Sun. 2009. Scaffolding and web concordancers as support for language learning. *Computer Assisted Language Learning* 22(4). 283–302. https://doi.org/10.1080/09588220903184518.

Charles, Maggie. 2007. Reconciling top-down and bottom-up approaches to graduate writing: Using a corpus to teach rhetorical functions. *Journal of English for Academic Purposes* 6(4). 289–302. https://doi.org/10.1016/j.jeap.2007.09.009.

Charles, Maggie. 2012. "Proper vocabulary and juicy collocations": EAP students evaluate do-it-yourself corpus-building. *English for Specific Purposes* 31(2). 93–102. https://doi.org/10.1016/j.esp.2011.12.003.

Chen, Hao-Jan Howard. 2011. Developing and evaluating a web-based collocation retrieval tool for EFL students and teachers. *Computer Assisted Language Learning* 24(1). 59–76. https://doi.org/10.1080/09588221.2010.526945.

Cheng, Winnie. 2011. *Exploring Corpus Linguistics*. Routledge.

Chinnery, George. 2008. ON THE NET You've Got Some GALL: Google-Assisted Language Learning. 12(1). 3–11. (4 March, 2024).

Conza-Armijos, Hover Ismael & Liliana Fernanda Celi-Celi. 2023. Data-Driven Learning in an EFL class: a study of Ecuadorian learners' perceptions. *INNOVA Research Journal*. Universidad Internacional del Ecuador 8(3). 37–50. https://doi.org/10.33890/innova.v8.n3.2023.2297.

Corino, Elisa. 2009. Data-driven Learning: tra lingue straniere e CLIL, tra ricerca e didattica. 8.

Corino, Elisa & Cristina Onesti. 2019. Data-driven learning: A scaffolding methodology for CLIL and LSP teaching and learning. *Frontiers in Education*. Frontiers Media S.A. 4. https://doi.org/10.3389/feduc.2019.00007.

Costas Gabrielatos. 2005a. Corpora and Language teaching: Just a fling, or Wedding bells? *TESL-EJ* 8(4). 1–37. (3 March, 2024).

Crosthwaite, Peter & Vit Baisa. 2023. Generative AI and the end of corpus-assisted data-driven learning? Not so fast! *Applied Corpus Linguistics*. Elsevier Inc. 3(3). https://doi.org/10.1016/j.acorp.2023.100066.

Crosthwaite, Peter, Alicia Gazmuri Sanhueza & Martin Schweinberger. 2021. Training disciplinary genre awareness through blended learning: An exploration into EAP students' perceptions of online annotation of genres across disciplines. *Journal of English for Academic Purposes* 53. 101021. https://doi.org/10.1016/j.jeap.2021.101021.

CSU Global. 2021. How Does AI Actually Work? *Colorado State University Global*. https://csuglobal.edu/blog/how-does-ai-actually-work#:~:text=AI%20systems%20work%20by%20combining.

Davies, Mark. 2009. Design, architecture, and Linguistic Insights. *International Journal of Corpus Linguistics* 14(2). 159–190. https://doi.org/10.1075/ijcl.14.2.02dav.

Doughty, Catherine & Jessica Williams. 1998. *Focus on Form in Classroom Second Language Acquisition*. New York: Cambridge University Press.

ffifuer Xl. *Esnd ffiffi?mrr-ilfrn lntenrational Confersnffi 2*15 ENGLISH DEPARTMENT FACULTY OF LETTERS AND CULTURE IN COLLABORATION WITH POST GRADUATE STUDY PROGRAM*.

Gabrielatos, Costas. 2005b. *Corpora and Language Teaching: Just a fling or wedding bells? **. (5 March, 2024).

Gocen, Ahmet & Fatih Aydemir. 2020. Artificial Intelligence in Education and Schools. *Research on Education and Media* 12(1). 13–21. https://doi.org/10.2478/rem-2020-0003.

Gordani, Yahya. 2013. The effect of the integration of corpora in reading comprehension classrooms on English as a Foreign Language learners' vocabulary development. *Computer Assisted Language Learning* 26(5). 430–445. https://doi.org/10.1080/09588221.2012.685078.

Granger, Sylviane & Gaëtanelle Gilquin. 2010. How can data-driven learning be used in language teaching? https://www.researchgate.net/publication/228984095.

Hadley, Gregory. 2002. An Introduction To Data-Driven Learning. *RELC Journal* 33(2). 99–124. https://doi.org/10.1177/003368820203300205.

Hadley, Gregory & Maggie Charles. 1094. Enhancing extensive reading with data-driven learning. *Language Learning & Technology ISSN* 21. 131–152. http://llt.msu.edu/issues/october2017/hadleycharles.pdf.

Haristiani, Nuria. 2019. Artificial Intelligence (AI) Chatbot as Language Learning Medium: An inquiry. *Journal of Physics: Conference Series* 1387(1). 012020. https://doi.org/10.1088/1742-6596/1387/1/012020.

Hartle, Sharon. 2023. From learner corpus to data-driven learning (DDL): Improving lexical usage in academic writing. *EuroAmerican Journal of Applied Linguistics and Languages*. E-JournALL 10(2). 9–31. https://doi.org/10.21283/2376905x.1.10.2.2748.

Haseski, Halil Ibrahim. 2019. What Do Turkish Pre-Service Teachers Think About Artificial Intelligence? *International Journal of Computer Science Education in Schools* 3(2). 3–23. https://doi.org/10.21585/ijcses.v3i2.55.

Hirata, Yoko & Paul Thompson. 2022. Communicative data-driven learning: A two-year pilot study. *ELT Journal*. Oxford University Press 76(3). 356–366. https://doi.org/10.1093/elt/ccab066.

Hsing Chin Lee. 2011. In defense of concordancing: An application of data-driven learning in taiwan. In *Procedia - Social and Behavioral Sciences*, vol. 12, 399–408. Elsevier Ltd. https://doi.org/10.1016/j.sbspro.2011.02.049.

Indra Nugraha, Sidik, Fauzi Miftakh & Kelik Wachyudi. 2017. *Teaching grammar through data-driven learning (DDL) approach*.

Istrate, Ana-Mihaela. 2018. Conference proceedings of»eLearning and Software for Education“(eLSE) Conference proceedings of”eLearning and Software for Education«(eLSE) Title: Artificial Intelligence and Machine Learning -Future Trends in Teaching ESL and ESP Artificial Intelligence and Machine Learning

-Future Trends in Teaching ESL and ESP. https://doi.org/10.12753/2066-026X-18-137.

Johns, Tim. 1986. *MICROCONCORD: A LANGUAGE LEARNER'S RESEARCH TOOL*.

Johns, Tim. 1991. *Should You Be persuaded: Two Examples of data-driven learning.* (Ed.) Tim Johns & Philip King. Birmingham.

Johns, Tim. 1997. Contexts: the Background, Development and Trialling of a Concordance-based CALL Program. 100–115. https://doi.org/10.4324/9781315842677-9.

Joshua. *Building a data-driven education system in the united states*.

Kiger, Patrick J. 1970. How Does AI Work? *HowStuffWorks*. https://science.howstuffworks.com/artificial-intelligence.htm#:~:text=AI%20systems%20work%20by%20combining.

Kilgarriff, Adam & Gregory Grefenstette. 2003. Introduction to the Special Issue on the Web as Corpus. *Computational Linguistics* 29(3). 333–347. https://doi.org/10.1162/089120103322711569.

Kotamjani, Sedigheh, Ommehoney Fazel, Habsah Hussin, Sedigheh Shakib Kotamjani & Shakib Kotamjani. 2017. English Language Teaching. 10(9). https://doi.org/10.5539/elt.v10n9p61.

Leńko-Szymańska, Agnieszka. 2017. Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology ISSN* 21(3). 217–241. http://llt.msu.edu/issues/october2017/lenko-szymanska.pdf.

Liang, Ling, Kai-ying Chen, Shu-yi Huang & Zhong-zheng Guo. 2023. A study on the application of a corpus-based data-driven learning method utilizing an online and offline blended teaching model in a college english reading course. In *Proceedings of the 2022 3rd International Conference on Big Data and Informatization Education (ICBDIE 2022)*, 81–101. Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-034-3_11.

Lin, M. H. & J.-Y. Lee. 2015. Data-driven learning: changing the teaching of grammar in EFL classes. *ELT Journal* 69(3). 264–274. https://doi.org/10.1093/elt/ccv010.

Lin, Phoebe. 2023. ChatGPT: Friend or Foe (to Corpus linguists)? *Applied Corpus Linguistics* 3(3). 100065. https://doi.org/10.1016/j.acorp.2023.100065.

Liu, MingYang. 2023. Exploring the application of artificial intelligence in foreign language teaching: Challenges and future development. *SHS Web of Conferences*. EDP Sciences 168. 03025. https://doi.org/10.1051/shsconf/202316803025.

Lüdeling, Anke & Merja Kytö. 2008. *Corpus Linguistics*. De Gruyter Mouton.

Luo, Qinqin. 2016. The effects of data-driven learning activities on EFL learners' writing development. *SpringerPlus*. SpringerOpen 5(1). https://doi.org/10.1186/s40064-016-2935-5.

Ma, Qing, Rui (Eric) Yuan, Lok Ming Eric Cheung & Jing Yang. 2022. Teacher paths for developing corpus-based language pedagogy: a case study. *Computer Assisted Language Learning* 1–32. https://doi.org/10.1080/09588221.2022.2040537.

Manyika, James, Michael Chui, Mehdi Miremadi, Jacques Bughin, Katy George, Paul Willmott & Martin Dewhurst. 2017. *A Future That Works: Automation, Employment, and Productivity*. McKinsey&Company. (25 March, 2024).

Marr, Bernard. 2023. 15 Amazing Real-World Applications Of AI Everyone Should Know About. *Forbes*. https://www.forbes.com/sites/bernardmarr/2023/05/10/15-amazing-real-world-applications-of-ai-everyone-should-know-about/?sh=7e7bfa5085e8. (25 March, 2024).

Meunier, Fanny. 2020. Data-driven learning: From classroom scaffolding to sustainable practices. *EL.LE*. Edizioni Ca Foscari (2). https://doi.org/10.30687/elle/2280-6792/2019/02/010.

Mizumoto, Atsushi & Kiyomi Chujo. 2015. 「論文」A meta-analysis of data-driven learning approach in the japanese EFL classroom 1. http://search.proquest.com.

Mizumoto, Atsushi & Kiyomi Chujo. 2016. Who is data-driven learning for? Challenging the monolithic view of its relationship with learning styles. *System*. Elsevier Ltd 61. 55–64. https://doi.org/10.1016/j.system.2016.07.010.

Mozelius, Peter & Niklas Humble. 2019. Artificial Intelligence in Education -a Promise, a Threat or a Hype? Artificial Intelligence in Education -a Promise, a Threat or a Hype? https://doi.org/10.34190/ECIAIR.19.005.

Mukherjee, Joybrato. 2004. Bridging the Gap between Applied Corpus Linguistics and the Reality of English Language Teaching in Germany. *Semantic Scholar*. https://doi.org/10.1163/9789004333772_014.

O'Sullivan, Íde. 2007. Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL* 19(3). 269–286. https://doi.org/10.1017/s095834400700033x.

Oktavianti, Ikmi Nur. 2015. DATA-DRIVEN LEARNING IN THE CLASSROOM: THE USE OF BRITISH NATIONAL CORPUS IN TEACHING VOCABOLARY. In *The 62nd TEFLIN Intenational Conference 2015*, 55–63. Udayana University Press.

OpenAi. 2022. ChatGPT . *chatgpt.com*. https://chatgpt.com.

Papaioannou, Vasiliki. 2018. *Teaching English as a Foreign Language through a Data-driven Learning perspective-using an annotated pedagogic corpus of English textbooks in a Greek high school class*.

Pedró, Francesc. 2019. *Working Papers on Education Policy 07 Artificial Intelligence in Education: Challenges and Opportunities for Sustainable Development Education Sector United Nations Educational, Scientific and Cultural Organization*. (25 March, 2024).

Pérez-Paredes, Pascual & Jose M. Alcaraz-Calero. 2009. Developing annotation solutions for online Data Driven Learning. *ReCALL* 21(1). 55–75. https://doi.org/10.1017/s0958344009000093.

Pérez-Paredes, Pascual, María Sánchez-Tornel, Jose María Alcaraz Calero & Pilar Aguado Jiménez. 2011. Tracking learners' actual uses of corpora: guided vs non-guided corpus consultation. *Computer Assisted Language Learning* 24(3). 233–253. https://doi.org/10.1080/09588221.2010.539978.

Rasikawati, Ira. 2020. Corpus-based data-driven learning in reading english for corpus-based data-driven learning in reading english for academic purposes academic purposes. https://digitalcommons.spu.edu/soe_etd/55.

Schaeffer-Lacroix, Eva. 2019. Barriers to trainee teachers' corpus use. In Peter Crosthwaite (ed.),. Routledge. https://hal.science/hal-02150903.

SCHMIDT, R. W. 1990. The Role of Consciousness in Second Language Learning. *Applied Linguistics* 11(2). 129–158.

Shen, Yiqiu, Laura Heacock, Jonathan Elias, Keith D. Hentel, Beatriu Reig, George Shih & Linda Moy. 2023. ChatGPT and Other Large Language Models Are Double-edged Swords. *Radiology* 307(2). https://doi.org/10.1148/radiol.230163.

Sihem Amer-Yahia. 2022. Towards AI-Powered data-driven education. In *Proceedings of the VLDB Endowment*, vol. 15, 3798–3806. VLDB Endowment. https://doi.org/10.14778/3554821.3554900.

Sinclair , Stéfan & Geoffrey Rockwell. 2016. Voyant Tools. *voyant-tools.org*. https://voyant-tools.org/.

Smart, Jonathan. 2014. The role of guided induction in paper-based data-driven learning. *ReCALL*. Cambridge University Press 26(2). 184–201. https://doi.org/10.1017/S0958344014000081.

Smith, Chris, Brian McGuire, Ting Huang & Gary Yang. 2006. *The History of Artificial Intelligence*. (24 March, 2024).

St, Elke & John. 2001. A CASE FOR USING A PARALLEL CORPUS AND CONCORDANCER FOR BEGINNERS OF A FOREIGN LANGUAGE. 5. 185–203. (5 March, 2024).

Sun, Yu-Chih & Li-Yuch Wang. 2003. Concordancers in the EFL Classroom: Cognitive Approaches and Collocation Difficulty. *Computer Assisted Language Learning* 16(1). 83–94. https://doi.org/10.1076/call.16.1.83.15528.

Sweller, John, Paul Ayres & Slava Kalyuga. 2011. *Cognitive Load Theory*. New York: Springer Science+Business Media, Llc, Cop.

Taylor, John R. 2012. *The Mental Corpus : How Language Is Represented in the Mind*. Oxford: Oxford University Press.

Thurstun, Jennifer & Christopher N Candlin. 1997. *Exploring academic English : a workbook for student essay writing*. Sydney National Centre For English Language Teaching And Research, Macquarie Univ.

Timms, Michael J. 2016. Letting Artificial Intelligence in Education Out of the Box: Educational Cobots and Smart Classrooms. *International Journal of Artificial Intelligence in Education* 26(2). 701–712. https://doi.org/10.1007/s40593-016-0095-y.

Tomasello, Michael. 2005. *Constructing a Language*. Harvard University Press.

Tribble, Christopher. 2014. *Textual Patterns: Key Words and Corpus Analysis In Language Education*. (7 March, 2024).

Wartman, Steven A. & C. Donald Combs. 2018. Medical Education Must Move From the Information Age to the Age of Artificial Intelligence. *Academic Medicine* 93(8). 1107–1109. https://doi.org/10.1097/acm.0000000000002044.

Wu, Xueqing & Rui Li. 2024. Effects of Robot-Assisted Language Learning on English-as-a-Foreign-Language Skill Development. *Journal of Educational Computing Research*. https://doi.org/10.1177/07356331231226171.

Xue, Liya. 2021. Using data-driven learning activities to improve lexical awareness in intermediate EFL learners. *Cogent Education*. Taylor and Francis Ltd. 8(1). https://doi.org/10.1080/2331186X.2021.1996867.

Yamamoto, Masahiro. 2023. *Evolution of Natural Language Processing Technology: Not Just Language Processing Towards General-Purpose AI Evolution of natural language processing technology: From "language" processing to general-purpose AI. Recent trends from a Japanese point of view*. (25 March, 2024).

Yang, Liu Ming. 2023. Application and Research on Foreign Language Teaching in the Context of Digital Transformation. (Ed.) S. Cheo-Chun & A.M. Belém Nunes. *SHS Web of Conferences* 159. 01011. https://doi.org/10.1051/shsconf/202315901011.

Yoon, Choongil. 2011. Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes* 10(3). 130–139. https://doi.org/10.1016/j.jeap.2011.03.003.

Yoon, Hyunsook. 2014. DIRECT AND INDIRECT ACCESS TO CORPORA: AN EXPLORATORY CASE STUDY COMPARING STUDENTS' ERROR CORRECTION AND LEARNING STRATEGY USE IN L2 WRITING. *Language Learning & Technology* 18(1). 96–117. (7 March, 2024).

Zeide, Elana. Carolina law scholarship repository carolina law scholarship repository the structural consequences of big data-driven education the structural consequences of big data-driven education the structural consequences of big data-driven education. https://scholarship.law.unc.edu/aidr_collection.

Zhang, Jie. 2022. Data-driven learning teaching model of college english based on mega data analysis. *Scientific Programming*. Hindawi Limited 2022. https://doi.org/10.1155/2022/3490594.

2017. *Artificial Intelligence Research, Development and Regulation*. IEEE-USA Board of Directors. https://globalpolicy.ieee.org/wp-content/uploads/2017/10/IEEE17003.pdf. (24 March, 2024).

2022. *AI BILL OF RIGHTS MAKING AUTOMATED SYSTEMS WORK FOR THE AMERICAN PEOPLE*. https://www.whitehouse.gov/ostp/aibillofrights.

Riassunto in Italiano

Questa tesi ha lo scopo di comparare degli esercizi realizzati attraverso strumenti di data-driven learning ed esercizi generati dall'intelligenza artificiale per determinare quale fra loro è il più adatto e il migliore nell'apprendimento della lingua Inglese. Attraverso questa scoperta, le metodologie utilizzate nell'insegnamento delle lingue potranno cambiare ed essere sempre più efficaci e innovative. La seguente tesi verrà suddivisa in tre capitoli, nei quali verranno trattati i temi principali di questa ricerca. Nel primo capitolo mi soffermerò sul data-driven learning e vedremo che cosa è, come viene utilizzato e quali sono i principali benefici e svantaggi. Nel secondo capitolo invece, metterò al centro dell'attenzione l'intelligenza artificiale, la sua storia, come funziona e le sue potenzialità. Infine, nel terzo capito mi concentrerò nell'effettiva comparazione di esercizi generati con l'AI ed esercizi realizzati con strumenti di data-driven learning per una ipotetica classe di studenti che studiano Inglese come seconda lingua e che possiedono un livello medio-basso di lingua Inglese. Attraverso i data raccolti, discuterò successivamente delle somiglianze, differenze, limitazioni e possibili miglioramenti. Al termine di questa comparazione condividerò il mio pensiero su ciò che studi futuri necessiteranno fare per poter approfondire maggiormente questo tema.

Il termine data-driven learning, abbreviato anche in DDL, fu coniato per la prima volta da Tim Johns verso la fine degli anni 80 e l'inizio degli anni 90 (Johns 1986). Secondo Liang et al. (2023:81) questo termine indica un metodo di insegnamento di una lingua straniera basato su dati raccolti in un corpus, permettendo di apprendere nozioni linguistiche in maniera diversa. Essenzialmente, gli studenti che studiano una lingua straniera con questa metodologia, assumono il ruolo di ricercatori perché navigano e osservano esempi linguistici all'interno di corpus, mentre gli insegnati assumono il ruolo di guida e assistenza nell'uso e nell'insegnamento di strumenti e applicazioni di DDL (Johns 1991). Con il termine "corpus" si intende un insieme di testi scelti accuratamente e raggruppati per specifici motivi (Cheng 2011:3). I testi presenti al suo interno possono provenire da fonti scritte e orali ed hanno la caratteristica di rappresentare un uso naturale e nativo di una specifica lingua (Zhang 2022:3; Costas Gabrielatos 2005:2). Al giorno d'oggi, è diventato sempre più facile accedere autonomamente, e in alcuni casi gratuitamente, a corpus elettronici attraverso software e applicazioni (Rasikawati 2020:6). Il DDL può essere diretto o indiretto, a seconda dell'approccio che l'insegnante sceglie di adottare con i propri studenti (Gabrielatos 2005b:1-37). Con un approccio diretto gli studenti dovranno autonomamente accedere al corpus elettronico e svolgere una ricerca necessaria per completare l'attività richiesta, mentre con un approccio indiretto, gli studenti verranno forniti di fotocopie con direttamente i dati necessari per completare l'attività (Corino & Onesti 2019:3-4). Sfortunatamente, entrambi hanno vantaggi e svantaggi. I benefici più importanti da citare sono l'autonomia e il lifelong learning per l'approccio diretto, mentre per l'approccio indiretto gli studenti possono conservare e consultare facilmente le fotocopie ricevute in classe e inoltre le attività risultano essere più adatte e ottimali per certi studenti. Gli svantaggi dell'approccio diretto sono le ore usate dagli insegnati per insegnare a utilizzare gli strumenti DDL e più in generale la necessità di fornire agli studenti computer per svolgere l'attività (Boulton 2017; Boulton 2012). Per quanto riguarda gli svantaggi dell'approccio indiretto, la più grande limitazione è il fatto che lo studente non sarà in grado di navigare autonomamente in un corpus (Boulton 2017).

Il termine "intelligenza artificiale" venne utilizzato per la prima volta nel 1945 da John McCarthy e oggi esprime l'abilità di macchine e computer di replicare una mente umana in grado di svolgere diversi compiti nello stesso modo in cui li svolgerebbe un essere umano (Anyoha 2017; Wartman e Combs 2018:1107). Il modo in cui l'intelligenza artificiale funzione è molto complicato. Secondo Kiger (1970), attraverso una combinazione di dati e algoritmi l'intelligenza artificiale impara da schemi e da dati raccolti. Per le sue infinite capacità, l'intelligenza artificiale può essere implementata in diversi settori tra cui anche l'istruzione (Borana 2016; Marr 2023). Liu (2023:1-3) in una sua ricerca, ha scritto in una lista alcuni utilizzi dell'intelligenza artificiale facendo riferimento all'insegnamento di una lingua straniera, e tra essi vengono riportati "speech recognition technology, machine translation technology, natural language processing technology and chatbot technology". Tutte queste tecnologie risultano importanti e molto utili, ma come è affermato dallo stesso Liu (2023: 3), queste tecnologie non potranno completamente sostituire insegnanti umani soprattutto per quanto riguarda aree importanti come la comunicazione interpersonale e l'espressività linguistica. Secondo una ricerca condotta da Crosthwaite e Baisa (2023:2), il DDL è in pericolo di venir surclassato dall'intelligenza artificiale. In quella ricerca, gli stessi Crosthwaite e Baisa, oltre a soffermarsi sui vantaggi degli strumenti di DDL, analizza anche cosa sarebbe necessario cambiare e modificare per fare in modo che quest' ultimi non vengano rimpiazzati dall'intelligenza artificiale. Per concludere, nello studio condotto da Liu (2023:4-5), risulta evidente come l'AI in futuro avrà un ruolo importante nell'insegnamento delle lingue straniere. Nonostante ciò, è fondamentale tenere in considerazione entrambi i benefici e i problemi che potranno insorgere.

In questo ultimo capitolo della tesi farò la comparazione tra due coppie di esercizi. Ciò, aiuterà a determinare quale tra DDL e AI è il migliore per l'apprendimento per studenti che studiano Inglese cose seconda lingua. Non avendo studenti reali a cui poter far testare gli esercizi, mi focalizzerò semplicemente nel paragonarli, analizzarli e descriverli. Nonostante questo, ho deciso di creare uno scenario ipotetico per una classe di ESL di 13 anni con un livello di lingua Inglese A2. Da una parte, gli esercizi creati con l'AI saranno generati da ChatGPT versione 3.5 con lunghi e dettagliate istruzioni che darò io, dall'altra parte, gli esercizi DDL che realizzerò saranno creati con l'aiuto di Sketch Engine con un approccio indiretto. Gli esercizi presentati e comparati riguardano principalmente i nomi numerabili e non numerabili. Il primo esercizio DDL presenta alcune linee di concordanza ricavate da un corpus con "some" e "any" evidenziati, e subito dopo le affermazioni in cui gli studenti devono indicare se sono vere o false. Gli studenti, basandosi sulle linee di concordanza e i dati che possiedono, devo individuare modelli e schemi per capire come e quando utilizzare "some" e "any" e completare l'esercizio. Anche l'esercizio generato con ChatGPT per essere risolto richiede che gli studenti analizzino i dati a loro messi a disposizione. ChatGPT è riuscito a realizzare alcune linee di concordanza con "some" e "any" evidenziati e successivamente ha realizzato un esercizio con alcune affermazioni che gli studenti devono completare. Passando al secondo esercizio, l'attività con il DDL si presenta con linee di concordanza ricavate da un corpus con "much" e "many" evidenziati, e subito dopo le frasi ci sono alcune indicazioni che lo studente deve svolgere. Prima di tutto, è stato chiesto agli studenti di trascrivere i nomi che seguono "much" e "many" nelle rispettive caselle e successivamente, con i dati raccolti dovranno completare alcune affermazioni scegliendo l'opzione corretta. Il secondo esercizio generato con ChatGPT nuovamente riesce a replicare una attività simile a quella dell'esercizio DDL. In questo esercizio, gli studenti, dopo aver letto ed esaminato delle linee di concordanza generate dall'AI, devono cerchiare i nomi numerabili e sottolineare i nomi numerabili presenti subito dopo "much" e "many" e successivamente completare un breve esercizio vero e falso.

Ciò che è possibile notare tra questi esercizi è principalmente come vengono strutturate e presentate le attività. Mentre gli esercizi DDL sfruttano linee di concordanza ricavate da un corpus, le frasi generate da ChatGPT sembrano essere simili e ripetitive fra di loro. Dai risultati ottenuti in questa ricerca è possibile notare vari punti di forza e varie limitazioni di entrambi gli approcci metodologici. Sfortunatamente, ciò non è bastato per riuscire a determinare quale tra DDL e AI è in grado di realizzare esercizi più efficaci per l'apprendimento della lingua Inglese. Ciononostante, mi auguro che futuri studi approfondiranno questa tematica testando ed esaminando esercizi di data-driven learning ed esercizi generai con l'intelligenza artificiale con due vere classi di studenti.