Review research paper

PLANNING AND DEVELOPING AN ONLINE ENGLISH WRITING COURSE FOR COMPUTER SCIENCE: A PRACTICAL GUIDE

Paul Hobbs-Koch, Paul Gahman, Maryna Rebenko

Friedrich-Alexander University, Erlangen, Germany and Taras Shevchenko National University of Kyiv, Ukraine

Abstract. *An online English course for computer science students at the B2 level was conceptualized for the E-learning institution Virtuelle Hochschule Bayern between Friedrich-Alexander University Erlangen-Nuremberg and the Technische Hochschule Nuremberg Georg Simon Ohm. The course consists of four separate modules that cumulatively target writing and language skills relevant to computer science students. Individual assessments at the end of each module serve as a review. Further, students write keystone pieces throughout the course that constitute a writing portfolio. The students' written pieces are targeted assignments that encapsulate both the writing and the language skills addressed in each module. For instructors, these provide insight into how well students acquired the requisite skills. The writing skills cover online writing tools (COCA, AntConc, OneLook, Quillbot), software documentation, expository text based on pseudocode, and finally, summarizing/paraphrasing tools. These are reinforced through language skills that either have greater prominence in the computer science field or are frequently used for the respective writing skills. Comma and hyphen rules, maintaining objectivity in writing, describing data in figures, and embedded clauses in English are some of the language skills covered. For materials development, a combination of corpus and AI writing tools were implemented to analyse and modify authentic scientific literature. Examples, keywords, and patterns of writing taken from authentic texts were modified to create exercises relevant to topics presented in the modules. The course provides strong didactic support for instructors whose aim is to facilitate a more natural set of writing tasks and students' ability to apply the learned skills beyond the classroom.*

Key words: *B2 online writing course, computer science and IT, writing and language skills development*

1. INTRODUCTION

Challenges and breakthroughs of the 21st century have incorporated a new concept in education – the application of technology. Since technology settings in education are constantly being updated, curricula, teaching methodology and learning tools can hardly stay uniform anymore. The latter has developed even more drastically as a result of the outbreak of the Covid-19 pandemic. The Corona-induced lockdown severely affected learning environments and thus, became a considerable challenge within education all over the world. On the other hand, remote education enhanced digital-literacy development for both students and teachers. In post-Covid Europe, one of the technological responses to pandemic-influenced learning has been online course platforms. Such initiatives have transformed conventional teaching methodologies, empowered teachers' digital expertise, and enabled students' more productive leverage of learning materials. All in all, the implementation of online courses reflects the growing global acceptance of online teaching and learning (Bashir et al., 2021).

With this in consideration, the authors of the paper aimed to conceptualize an online writing course on request of The Virtuelle Hochschule Bayern, the nation-wide e-learning platform in Germany. Specifically, the authors planned and designed the materials for developing computer science and IT students' effective writing and vocabulary skills at the B2 level (Common European Framework of Reference).

2. LITERATURE REVIEW

Online education's exponential ubiquity has become a reality that continues to generate challenges for both instructors and learners. These include an online course development, delivery, testing, and evaluation (Smith & Ragan, 1999), which, according to Muffoletto, might be "an adventure among instructors and students" (1997, p. 50). For Savenye et al. (2001), such "an adventure" resulted in working out systematic principles of instructional design for online courses, which encompassed a learners' analysis; a learning objective; methods, media and materials selection; media and materials use and structuring; implementation of new approaches; a learner participation requirement, and instruction revision and evaluation. Significant was the development of the guidelines on online writing instructional design. Moreover, Savenye et al. determined the key attributes and best practices of on-student and faculty support for fostering learning online and reaching more numerous and diverse learners through technology integration (2001, p. 373–378).

The overall integration of modern technology into academic settings resulted in the development of interaction tools that meet learners' need for efficient online courses (Savenye, 2005). With Koehler and Mishra's (2008) technological pedagogical content knowledge model (TPCK), Tai et al. (2015) worked out an online English writing course for students in the medical field. The researchers focused primarily on writing skills as a crucial element of language learning and carried out a complete study, starting from the course development to the evaluation of outcomes through students' learning progress and feedback.

Tai et al.'s teaching strategy included TPACK – a simplified Koehler and Mishra's (2008) framework made for the sake of remembering – a three-facet component model: technology, pedagogy, and content. Technology was introduced on e-campus online writing teaching and learning platform, which combined learning interaction, evaluation, and information management and was maintained by Chang Gung University of Science and Technology (Taiwan). Pedagogy implied the use of the process approach to viewing writing as thinking and discovery rather than the traditional "product", which concentrates on the output and accuracy of writing (Hyland, 2002, p. 20–21). Finally, the content component targeted the development of English as a Foreign Language (EFL) writing skills. Each learner had to review their peers' initial writing drafts, and then give feedback by filling out a review form. Each student thus received three feedback forms and was tasked with modifying the writings according to the peer comments. Additionally, two blended modes of teacher's indirect and direct feedback were integrated into the TPACK model course. Significant is that in spite of a moderate satisfaction level toward the TPACK writing course among the participants (the majority found the writing instructions and peer feedback inconsistent), the course implementation resulted in learners' writing competence improvement, which was proved by the descriptive statistics, MANOVA and content analysis data (Tai et al., 2015, p. 785–786).

With concern about students' inequality in learning materials and instructors' inability to identify specific qualities of online learners, Borgman and Dockter (2018) showcased a user-centered online course approach as an efficient strategy to promote learning and decrease student attrition. The researchers followed Oswal and Meloncon's (2014) imperative on course design, technology appropriateness and accessibility for the students' successful orientation in learning. As a result, they considered student users as a key element of user-centered design approach and utilized them to make the online courses more universal. Borgman and Dockter's principles correlate with Womack's (2017) idea of teaching as accommodation whereby universal course design meets varied learners' needs and promotes inclusivity. Accordingly, the researchers instructed the tutors to pay closer attention to the learning materials' delivery methods. They also aimed to raise online course developers' awareness that aside from the course content there is an issue of the design possibilities and thus inclusivity.

Bielousova (2017) addressed the issue of designing and integrating English for Specific Purposes (ESP) online courses for B2 technology students within the project "Application of e-learning in foreign language teaching at the Faculty of Manufacturing Technologies". In her course design, the author followed skills and learning-centered approaches, i.e., analyzed target needs and the learning environment, worked out an adequate syllabus, then selected, adapted, and prepared corresponding learning materials. Considering that ESP authentic texts are much more abundant in language forms than the texts constructed for language teaching purposes (Gilmore, 2007, p. 101), the researcher developed a database of short authentic texts with illustrations and other visual features. The constructed interactive teaching materials were integrated into the university Moodle platform of the Technical University in Slovakia.

Another powerful approach within online language learning is corpus linguistics tools application. Bednarek and Carr (2020) introduced a way of performing digital

communications research without any programming knowledge through the computer-assisted analysis of newspaper articles. The researchers applied corpus linguistic software (WordSmith and AntConc) to prove its usefulness for the analysis of standard and atypical language, sources, stigma and responsibility, framing, and project-specific text analysis (2020, p. 10–13). Another study, Ebrahimi and Faghih (2017), investigated the effectivity of integrating corpus linguistics tools into online courses. The researchers analyzed 32 language teachers' evaluations of online educational courses in corpus linguistics. They concluded that to facilitate virtual learning environments it is necessary to provide technology in educational settings; incorporate corpus linguistic instructions in all stages of language teacher education degree programs; emphasize corpus linguistics' practical aspects in pedagogical science; introduce user-friendly tools and encourage indirect use of corpora in case of lack of technological facilities; provide effective instructional materials along with sufficient instructor support; and finally, motivate learners to reflect critically on the employed approach (2017, p. 134).

Chang and Kuo (2011) applied corpora, comprised of 60 computer science articles and genre analysis to develop teaching and research-supported materials tailored to computer science graduate students. A seven-stage study (corpus compilation, genre analysis, move tagging, text analysis, online materials development, and experimental teaching and assessment) resulted in the development of online materials and specific lexico-grammatic features (2011, p. 228–230). It was then incorporated into the English for Academic Purposes (EAP) course website, constructed on Moodle, and pilot-tested at two Taiwan universities. Supplemented by a web-based concordance populating tool, a collocation tool, and self-developed rhetorical moves and keywords, the corpus-based online EAP courseware was targeted to facilitate non-native students' writing process. Pre- and post-learning assessment showcased the learning effectiveness.

Based on the aforementioned discussion, we developed the online course with a strong focus on technology, pedagogy and content interaction. Motivating students to improve their written skills within technical contexts was the primary impetus behind course development. Providing students with the tools and skills necessary to produce written work commensurate with the university level and beyond further shaped the core tenants of the course. The pedagogical and methodological approach employed in this course represents one solution to the inherent challenges behind online education for both students and instructors.

3. PARTICIPANTS AND SETTING

The targeted group of participants is open to the students with a technical background all throughout Bavaria (Germany). Before the course, students are able to register if they have completed a B1 course or if they already have a B1 level certification. If the students are confident with the tasks, they will continue onto the first module, otherwise, they will be referred back to other B1 writing courses.

The course was designed in three steps. In early 2022-23 we analysed and modified authentic scientific literature. Four separate modules cumulatively targeted writing and language skills relevant to computer science students.

Module 1 "Writing Mechanics and Online Writing Tools" was meant to teach the students how to express and describe hypothetical situations using appropriate grammatical structures and vocabulary within the context of deepfake technology. Also, the students learned to use a variety of online corpus and writing tools which they might encounter while writing. These topics were introduced through the rules of commas, hyphens, common abbreviations and semicolons in technical English writing.

The content topic of Module 2, "Software Documentation and The Revising Process", aimed to teach students how to produce a basic software document with suitable structure and layout using appropriate language expressions; to use vocabulary appropriately in context; and how to draft, revise, and edit their writing while maintaining objectivity and improving accuracy. Above all, students learned when and where to employ active and passive voice in technical writing most appropriately.

Module 3 "Describing Data Analytics and Code" introduced students to various statistics, data analytics, and data types illustrated through line, pie and bar charts. The context of Digital Twin technologies, an amalgamation of physical and digital phenomena for data analytics, served as an example. The module also developed linguistic expressions for formulating and explaining pseudo and actual code. It fostered students' ability to explain abstract (pseudocode) and concrete (the meaning behind the actual code) phenomena through sign-posting devices to express the logical and structural development of writing.

"The Ethics behind Copying Code", Module 4, introduced the two scenarios of code – designed to be shared and copyrighted. It thereby launched the concept of plagiarism, steps students should take to avoid it in programming assignments, and the mechanism of citing another author's piece of work. The practice of paraphrasing and summarizing with and/or without QuillBot system tools together with various forms of subordinate, embedded and/or relative clauses enabled students to achieve greater syntactic diversity in their linguistic expression. A module quiz as an individual assessment tool at the end of each module served as a review.

Finally, we proofread the developed materials, worked out module quizzes' content and writing tasks for the students' portfolio, which in the end, would constitute the final grade. Also, we started the first round of the trial course in 2023 to test how the students receive the materials and course in general. The course assessment amounted to 40% quizzes (10% each) and 60% portfolio (15% each).

4. MATERIALS AND METHODS

4.1. Course Development

This section offers insight into the course conception and development, detailing the authors' transition from drawing language skills solely on the Common European Framework of Reference (CEFR) to identify key areas of writing that are particularly pertinent to computer scientists. Through careful consideration of these areas and the effective use of authentic materials, the course developers were able to provide targeted scaffolding for the presentation of specific writing skills as well as to develop more authentic exercises and testing materials throughout the course.

4.1.1. Use of CEFR for Identification of Language Skills

In the initial development of the course, the CEFR was consulted to extract which language students should achieve at the B2 level. Specifically, sections relevant to writing were investigated to ensure that students would be able to "[…] write clear, detailed texts on a variety of subjects related to his/her field of interest, synthesising and evaluating information and arguments from a number of sources" (Council of Europe, 2021, p. 66). Within the context of this course, the practical implication of these writing standards is that students must demonstrate the ability to produce well-structured texts in terms of what is relevant to their area of expertise, specifically in the field of computer science. Students would also have to be able to show accurate use of grammar relevant to the type of text they are producing based on the input received throughout the course itself.

The CEFR provides descriptions at each level and for individual registers that outline what users of a language *can do*. To this end, the *can dos* outlined in the CEFR, such as being able to use hypotheticals accurately (Council of Europe, 2021, p. 174), were carefully considered in the creation of exercises and practice that focus on providing students with the opportunity to utilize language and vocabulary relevant to their field of study. These exercises were designed to aid students in gaining a comprehensive understanding of how the language and vocabulary can be applied to achieve a specific goal, such as producing detailed software documentation or summarizing and paraphrasing existing texts for later use in scientific or academic writing.

4.1.2. Writing Skills for IT

Alongside the traditional descriptions provided by the CEFR, new descriptors in the area of mediation were considered. Mediation is "[r]elaying specific information, [e]xplaining data verbally, [p]rocessing text or [t]ranslating a written text for someone else" (Piccardo et al., 2019, p. 23). The capacity to choose and interpret key empirical information presented in the form of a graph is critical for written reports. In other words, students must be able to pick and analyse the most significant and relevant data elements that are visually presented in a graph (Council of Europe, 2021, p. 97). With this in mind, students at the B2 level should be able to accurately convey as well as interpret data found in a graph in written form. Not only is this a necessary skill when interpreting data, but it additionally plays an important role specifically in the presentation of results. According to Glazer, being able to interpret data and draw conclusions from this data in order to report key findings are necessary skills for all scientists at all levels (2011, p. 189–190).

Being able to explain code in the form of pseudocode is another necessary mediation skill that is covered in this particular course. Emphasis is placed on this skill to ensure that later in students' careers, they will be able to actively participate in exchanges between IT professionals and non-IT professionals working together in a business setting. As not everyone working in software development has knowledge of programming, it is necessary for students to be able to explain the functionality of code through pseudocode. Because the authors of the course could not be certain what the programming background of students is, it was decided to use pseudocode as a language-agnostic way of describing what code does. Due to the basic nature of the language in pseudocode, students would then be guaranteed to finish the course having learned useful language for a wide range of programming languages as well as software functionalities.

4.1.3. Writing Skills for Academic Settings

From an academic point of view, there were several writing skills that was found to be of importance for computer scientists. These skills were based on the ability to paraphrase other's writing. This is a necessary skill in that those working in IT often have to include information found from other sources in the various forms of documentation that they produce. Due to the frequent inclusion of information from external sources in various forms of documentation, paraphrasing is a useful skill for those currently in and wishing to stay in academia. One particular area that students learn to focus on in one of the modules is how to avoid ambiguity and subjectivity in writing.

In these chapters covering ambiguity and subjectivity, students focus on identifying various forms of vague language, reformulating long noun strings, as well as language that is ambiguous and subjective. For example, students are instructed they should avoid first person pronouns, vague terms such as *lose* or *tight*; verbs such as *think* or *believe*; conjunctions such as *while*, *when*, or *whenever*; and phrases such as *in my opinion* or *according to me*. Students are then asked to identify ambiguous and subjective language in the various exercises and to rewrite sentences making them more precise and less subjective.

In addition to working on making their writing more precise and less subjective, the authors of the course included additional work on avoiding long noun strings. Long noun strings in English are confusing and create confusion when reading because of how the words are syntactically grouped (Spärck Jones, 1983, p. 2). For example, the long noun string *sensor modulation activation process accuracy rate* is confusing due to not immediately being able to understand how the words syntactically interact. To reduce the ambiguity of the noun string, students are given examples of how such strings should be broken down into more swallowable chunks of language embedded in prepositional phrases or possessive constructions. In doing so, the long noun string becomes: *accuracy rate of the activation process for sensor modulation* or *activation process' accuracy rate for sensor modulation*. These constructions are less ambiguous as they do not cause any confusion in syntactic processing.

4.2. Materials

This section will explain the reasons why authentic materials were integrated into the course. Additionally, it will cover the various text type characteristics and discuss the reasoning for including various forms of input throughout the course.

4.2.1. Authentic Materials

The authors designed this course with two primary objectives for students: acquiring knowledge relevant to the academic computer science field and developing transferable skills for professional working environments, particularly in business contexts. To achieve these goals, the incorporation of authentic materials was deemed essential.

According to Rogers (1988), language learners require adequate contact with authentic materials in order to gain the ability to communicate effectively. While modified texts have their place in a classroom, authentic texts offer students valuable insight into real-world contexts where they may need to function at a high level. As stated by Akbari and Razavi, "by using authentic materials students know how to use [the] language in [the] real world and improve their proficiency level in four skills of language learning" (2015, p. 108). Thus,

the authors integrated authentic materials related to computer science into the course to enhance students' learning experience.

Some examples of authentic materials that were incorporated into the course included the use of pseudo code and the language used to describe it. It is often the case that in a professional environment, individuals working in IT or computer science have to communicate what they are doing to managers or others from other departments what they are doing and why. Often this is done in the form of a presentation but can also be done in reports. Having considered this, the authors decided it would be good to teach students how to interpret code in a way that would be more easily understood by lay-people, hence, the mediation of pseudo code. Students were provided with authentic examples of pseudo code and detailed descriptions of the code so that they would be able to see and understand how the code could be mediated to individuals in a way that could be understood by those who do not work in the field.

Furthermore, the authors of this course analysed documents of pseudocode in order to extract useful phrases which students could then use in their own descriptions. Essentially, these phrases are bundles of words that students can use in their own writing without having to worry about plagiarism. Originally Biber *et al.* identified what they called "recurrent word combinations" or "clusters" (1999, p. 990) as recurring sequences of words found within a register. Such *clusters* were referred to by Biber, Conrad, and Cortes (2004) in their publication in which they investigated similar phenomena of multi-word sequences in university teaching and textbooks. In their conclusion, the researchers state, "[t]hese sequences of words can be regarded as structural 'frames', followed by a 'slot'" (Biber et al., 2004, p. 399). The authors concluded that multi-word sequences, when used in textual contexts, serve as significant constituents of discourse and are closely linked with fundamental communicative roles. Therefore, the authors of this course concluded that these bundles are exceptionally beneficial for students in that students no longer have to come up with their own formulation. By providing them, students are able to 'plug and play' so to speak in that they only have to choose the correct bundle or phrase in order to convey particular information. This frees up students' minds in that they can simply commit the phrase to memory and never have to worry about the individual functions of the words making up the phrase.

4.2.2. Secondary Input through Authentic Materials

Authentic materials also provide students with secondary input throughout the course, which further supports their language development. According to Swaffer (1985), not only can authentic texts to meet a particular learning goal, but they can additionally be used to help increase the frequency of vocabulary and grammar found within specific subject areas being addressed in a course, such as the language found in a software document. The authors chose resources carefully to improve students' comprehension of writing styles and language use in professional and academic contexts relevant to the field of computer science. For example, the authors of the course decided to include an academic text on Deepfake technology. In the article, the instances of hypothetical language were highlighted as a way of introducing the concept of how hypothetical language is used in such a piece of writing. Furthermore, hypotheticals show up again in pseudocode through the use of conditional phrases, thus reinforcing language introduced earlier in a different authentic context. Materials were most readily sourced from online computer science journals such as IEEE and the online corpora COCA and BNC.

4.2.3. Challenges and Solutions for Using Authentic Materials

Due to modern copyright law, it is and can be difficult to implement the use of authentic materials in a course. However, it should be noted that aspects of those copyrighted materials can be quite relevant and can be analysed by course developers to identify various patters of use in vocabulary and grammar. For example, because software documents tend to be very product specific and are not produced under a creative commons license, those materials can be analysed to identify patterns of vocabulary and grammar. This enabled the authors to produce a software document for one of the tools introduced to students (COCA) before the students have worked with the tool itself. This provided students not only with an example of a software document that they would then later use as a model for their own software document, but also introduced the tool to the students in a way that a lot of software they may encounter in the future while working in their careers. With regard to the appropriateness of the materials, minimal changes were made. Because research articles written by experts in the field formed the majority of the authentic materials, minor portions of text deemed inappropriate for students at the B2 were adjusted. Moreover, examples were designed to expose students to a variety of language features found in academic and scientific writing through analysis of authentic documents having used the tool, Ant-Conc. In the following sections, the ways in which the various corpus tools implemented to develop the course will be discussed and examples will be provided. Furthermore, the introduction of these corpus tools to students and their potential for improving students' writing abilities will be explored.

4.2.4. Text Type Characteristics

According to Ozverir, Osam, and Herrington, even though many students can talk about grammar rules, many students are poor at applying the language in real-world contexts (2017, p. 261). Authentic materials were additionally used to help support and to incorporate various linguistic features found within various types of texts from the field of computer science and IT. These texts included full academic texts as well as selections of texts, software documents, pseudocode, and graph descriptions. These materials not only provide students with the opportunity to gain useful vocabulary from sources that they may encounter later on in their careers, but also help to highlight specific types of linguistic features that students are then later drilled, quizzed, and tested on later.

In a paper published by Granger and Bestgen (2014), several studies investigating how both intermediate and advanced learners used both high and low frequency words in collocations. It was found that learners tend to overuse collocations containing high-frequency words and would avoid those containing low-frequency words. It was very clear to the authors of the English IT course that in order for students to advance and improve their writing skills, they had to ensure that students were identifying and using collocations found within authentic materials, such as software documents. In doing so, students would be "freed from the task of composing such sequences word-by-word" (Pawley & Syder, 1983, p. 208).

Within an academic scope, as shown in Biber et al. (1999, p. 606), non-finite relative clauses are found with a very high frequency, and as discussed by Rafajlovičová (2012, p. 15), reduced relative clauses are extremely common in the written register. As a result, the authors of the course included exercises testing students' abilities to transform relative clauses into reduced relative clauses as a means of ensuring that students would have varied sentence types as found in academic writing.

4.3. Corpus and Online Tools

This section will outline how the authors implemented the tools: AntConc, the Corpus of Contemporary American English, the British National Corpus, and iWeb. The tools were selected to create genuine examples that reflect how language and vocabulary are commonly used within the context of computer science. Lastly, this section sets out to detail how these tools were integrated throughout the course to provide students with opportunities to practice using them and to help improve the quality of their own writing.

4.3.1. AntConc

AntConc is "a freeware corpus analysis toolkit for concordancing and text analysis" (Anthony, 2022). AntConc allows users to input a text file or mini corpus for analysis and then search for specific words, phrases or patterns within that text. In developing the course, the authors gathered various pieces of writing such as in software documents and academic texts on subjects ranging from deepfakes to digital twin technologies, and even code.

These texts were examined to identify patterns in grammar and vocabulary. By analysing authentic texts in this manner, the course creators ensured that the materials used throughout the course mirrored the language and vocabulary relevant to the IT field, ultimately helping students develop the necessary competencies for success in their future careers in academia or industry. For example, the authors of the course gathered multiple software documents using the tool's word list and collocations feature. After generating a word list, the authors were able to identify those technical terms which were of high frequency. After doing so, the tool's collocations feature was used to identify collocations based on those high-frequency words.

Furthermore, AntConc was used to help create exercises for students that were based on these authentic materials, which would help students develop their language and writing skills in a way that resembles native speakers or those who have been working in the field, either in IT or academically. This was done so that the authors would not rely on their intuitions, but rather relied on data from texts produced by individuals already working in the field of IT. This is done because studies of native speaker intuitions provide unclear results. This means that just because a native speaker thinks something is the case regarding their language, does not mean that the specific language feature in question is as the native speaker says. In a study performed by Schmitt and Dunham, when asked to rank synonymous words in terms of overall frequency, even native speakers' rankings were inaccurate (1999, p. 407). Due to this disconnect between what native speakers think about their language and what native speakers actually do, the use of corpus tools for developing a course for specific purposes can be better understood.

Not only was AntConc valuable for the course creators, but there is also recognizable value for students. Moreover, AntConc was integrated into the course by providing students with guided instruction on utilizing the software to independently identify phrases, collocations, and word lists from materials they frequently encounter. This approach will hopefully empower students to take greater ownership of the language learning process while engaged in the course and into the future.

4.3.2. COCA, the BNC, and iWeb

O'Keefe, McCarthy and Carter additionally mentioned in their 2007 book *From Corpus to Classroom* that much of the information found within traditional language materials is based solely on authors' intuitions regarding the language rather than on data found within corpora. The authors of the book state, "students who encounter only scripted […] language have less opportunity to extend their linguistic repertoires in ways that prepare them for unforeseeable interactions outside the classroom" (2007, p. 21). For this reason, the authors of the course wanted to ensure that the language contained within in the course reflected what native speakers and regular users of English would use in those specific contexts by using the Corpus of Contemporary American English (COCA), the British National Corpus (BNC), and iWeb.

The Corpus of Contemporary American English (COCA) is a large, online corpus of American English. The corpus currently contains more than 1 billion words, making it an extremely large database of language (Davies, 2008). According to O'Keefe, McCarthy and Carter (2007, p. 4), any corpus over 1 million words in terms of its database size is to be considered large. COCA is designed to be a representative sample of English as its data stems from a wide range of contexts, including spoken language, fiction, popular magazines, newspapers, and academic texts. The use of AntConc in the previous section served mainly to identify those high-frequency lexical items found within authentic materials. For the purposes of developing this course, the authors mainly used COCA to check how those lexical items are used in context.

The British National Corpus (BNC) is a large corpus, albeit smaller than COCA, containing over 100 million words from various dialects of British English (BNC Consortium, 2019). The corpus itself functions in the same way as COCA and can be used to identify lexical use in context in a wide range of registers, e.g., spoken, fiction, magazines, newspapers, and academic.

Another tool found within the same group of online corpora is iWeb. iWeb is a collection of corpora containing over 14 billion words, which is approximately 14 times larger than that of COCA and is distributed across 22 million web pages. iWeb is related to numerous other English corpora created by the same individuals as COCA and the BNC providing insights into the English language. Unlike other large web-based corpora, the nearly 95,000 websites included in iWeb were specifically selected, each containing over 145,000 words (The iWeb Corpus).

Within the course itself, COCA, the BNC, and iWeb were used to show students how they can check the use of various vocabulary and how they collocate. The authors of the course aimed to ensure that students were able to take greater control of their language learning and show students how they can identify the use of key vocabulary across register, how similar words are used, and how to identify high frequency collocations. With this goal in mind, after completion of the course, students will have left knowing that if they are faced with challenges in writing regarding word use, they would then have the necessary knowledge of how to use these tools to complete specific written tasks using appropriate language.

4.3.3. OneLook Thesaurus

OneLook Thesaurus is a tool for students in that it has the unique ability to search for words based on a description or definition of a word. It can even search for words based

on specific letters or patterns and has the ability to find words that rhyme or match a certain length. For example, an individual can type into the search field 'words that rhyme with *time* meaning to make something ready for use or action' to find the verb *to prime*. Additionally, for language learners, OneLook can help identify synonyms and other related words to a term. This tool can be employed to enhance a learner's vocabulary and identify the most fitting word or phrase for a specific context, given its capacity to search for terms based on descriptions. For example, by entering the description 'cooked in the oven', a student would then better understand that the verb *cooked* would not be used in this context, but rather the verbs *baked*, *roasted*, or *grilled*. This functionality proves particularly beneficial during the writing process, especially when the student is unfamiliar with synonymous terms due to having limited vocabulary. Overall, the tool aids in the avoidance of overly simple language in both academic and scientific writing. Within the computer science course, the authors guide students through the various features of OneLook and provide students with exercises, helping them to gain a better understanding of its most helpful features, searching by word, by description, and by pattern.

4.3.4. Oxford Collocation Dictionary – freecollocation.com

For individuals who may struggle to consistently use corpus tools to identify collocations, the Oxford Collocation Dictionary is a simple alternative. The dictionary contains over 150,000 collocations for nearly 9,000 words (Online OXFORD Collocation Dictionary of English). This is a tool for those who wish to have a simple alternative to using corpus tools. However, one drawback is that the size of the dictionary is nowhere near the size of the corpora mentioned in previous sections and can often provide users with no answer to their query. Fortunately for such searches providing no results, students can refer back to those language corpora such as COCA, the BNC, or iWeb.

4.3.5. QuillBot

Lastly, the authors of the course introduced the writing tool, QuillBot. QuillBot is an AI-based writing tool that takes advantage of machine learning to rephrase, summarize, and paraphrase text. The tool uses natural language processing (NLP) to process and understand the context of the text provided to the AI, then suggests alternate word choices, grammatical structures, and other changes to improve the writing overall.

The tool can be used to paraphrase an entire piece of writing at one time, or simply a small portion of text such as a sentence or phrase. The tool also different modes, ranging from a standard mode designed to optimise the output for clarity and a fluency mode which aims to improve the readability of the text.

QuillBot is not designed to perform a writing task for a writer. Rather, it is designed to guide learners to improve their writing skills by helping them express their thoughts in a clearer and more concise manner, as well as for those individuals who wish to improve their command of written English through helping the writer identify synonymous vocabulary and alternative grammatical structures for expressing similar ideas. The tool does so by analysing the existing language in a text and then offering suggested changes based on the vocabulary and grammar already present in the writing. The tool is also valuable for writers who are required to paraphrase existing texts for academic or professional purposes, facilitating the avoidance of plagiarism. Therefore, the authors included QuillBot in the course for paraphrasing techniques and accompanying guided exercises.

5. CONCLUSION

The designed B2 online computer science writing course:

- meets the needs of students with a technical background;
- targets not teaching English to IT students, but improving their writing and language skills;
- facilitates a student-friendly learning environment;
- provides strong didactic support for instructors aimed to facilitate a more natural set of writing tasks;
- ensures students' ability to apply the learned skills beyond the classroom;
- and serves as constant student-teacher feedback for continued improvement.

Ultimately, the course development is an iterative process as nothing is ever written in stone. The underlying pedagogical principles of the course reflect contemporary advances in overcoming challenges to online instruction, from which students can concretely benefit. As a prospect for further research, the present course with corpus integration could ensure statistically indicating the degree of success and effectiveness of the methodology applied in the present course. In doing so, the didactic process outlined here can be objectively evaluated as a feasible guide for course development. Even if a course developer is not an expert in the field of the course they are developing, using an approach similar to what was outlined here in this paper can allow a teacher to produce customized and authentic material that is relevant to students.

REFERENCES

Akbari, Omid, and Azam Razavi. 2015. "Using Authentic Materials in the Foreign Language Classrooms: Teachers' Perspectives in EFL Classes." *International Journal of Research Studies in Education* 4 (5). https://doi.org/10.5861/ijrse.2015.1189.

Anthony, Laurence. 2022. AntConc (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software.

Bashir, Amreen, Shahreen Bashir, Karan Rana, Peter Lambert, and Ann Vernallis. 2021. "Post-COVID-19 Adaptations; the Shifts Towards Online Learning, Hybrid Course Delivery and the Implications for Biosciences Courses in the Higher Education Setting." *Front. Educ*. 6: 711619. https://doi.org/10.3389/feduc.2021.711619.

Bednarek, Monika, and Georgia Carr. 2020. "Computer-Assisted Digital Text Analysis for Journalism and Communications Research: Introducing Corpus Linguistic Techniques That Do Not Require Programming." *Media International Australia*, 1–21. https://doi.org/10.1177/1329878x20947124.

Biber, Douglas, Susan Conrad, and Viviana Cortes. 2004. "If You Look at ...: Lexical Bundles in University Teaching and Textbooks." *Applied Linguistics* 25 (3): 371–405. https://doi.org/10.1093/applin/25.3.371.

Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan, Randolph Quirk, and et al. 1999. *Longman Grammar of Spoken and Written English*. Longman.

Bielousova, Rimma. 2017. "Review of Developing Materials for English for Specific Purposes Online Course within the Blended Learning Concept." *TEM Journal* 6 (3): 637–42. https://doi:10.18421/TEM63-28.

BNC Consortium. 2019. "British National Corpus (BNC)." English-Corpora.org. https://www.english-corpora.org/bnc/

Borgman, Jessie, and Jason Dockter. 2018. "Considerations of Access and Design in the Online Writing Classroom." *Computers and Composition* 49: 94–105. https://doi.org/10.1016/j.compcom.2018.05.001.

Chang, Ching-Fen, and Chih-Hua Kuo. 2011. "A Corpus-Based Approach to Online Materials Development for Writing Research Articles." *English for Specific Purposes* 30 (3): 222–34. https://doi.org/10.1016/j.esp.2011.04.001.

Council of Europe. 2021. "Common European Framework of Reference for Languages: Learning, Teaching, Assessment." *Council of Europe*.

Davies, Mark. 2008. The Corpus of Contemporary American English (COCA): 1 billion words +, 1990-present. Available from https://www.english-corpora.org/coca/.

Ebrahimi, Alice, and Esmail Faghih. 2017. "Integrating Corpus Linguistics into Online Language Teacher Education Programs." *ReCALL* 29 (1): 120–35. https://doi.org/10.1017/s0958344016000070.

Gilmore, Alex. 2007. "Authentic Materials and Authenticity in Foreign Language Learning." *Language Teaching* 40 (2): 97-118.

Glazer, Nirit. 2011. "Challenges with Graph Interpretation: A Review of the Literature." *Studies in Science Education* 47 (2): 183–210. https://doi.org/10.1080/03057267.2011.605307.

Granger, Sylviane, and Yves Bestgen. 2014. "The Use of Collocations by Intermediate vs. Advanced Non-Native Writers: A Bigram-Based Study." International Review of Applied Linguistics in Language Teaching 52 (3). https://doi.org/10.1515/iral-2014-0011.

Hyland, Ken. 2002. *Teaching and Researching Writing*. Pearson, Harlow and London.

Koehler, Matthew J., and Punya Mishra. 2008. "Introducing TPCK." In: AACTE (Ed.), *Handbook of Technological Pedagogical Content Knowledge* (TPCK) for Educators. Routledge, New York, 3–29.

Muffoletto, Robert. 1997. "Reflections on Designing and Producing an Internet-Based Course." *TechTrends*, 50–3.

O'keeffe, Anne, Michael McCarthy, and Ronald Carter. 2007. *From Corpus to Classroom: Language Use and Language Teaching. Cambridge*; New York: Cambridge University Press.

"OneLook Thesaurus." n.d. Www.onelook.com. https://www.onelook.com/thesaurus/.

"Online OXFORD Collocation Dictionary of English." n.d. Www.freecollocation.com. https://www.freecollocation.com/.

Oswal, Sushil K., and Lisa Meloncon. 2014. "Paying Attention to Accessibility When Designing Online Courses in Technical and Professional Communication." *Journal of Business and Technical Communication* 28 (3): 271–300. https://doi.org/10.1177/1050651914524780.

Ozverir, Ildeniz, Ulker V. Osam, and Jan Herrington. 2017. "Investigating the Effects of Authentic Activities on Foreign Language Learning: A Design-Based Research Approach." *Journal of Educational Technology & Society* 20 (4): 261–74. http://www.jstor.org/stable/26229222.

Pawley, Andrew, and Frances Hodgetts Syder. 1983. "Two Puzzles for Linguistic Theory: Nativelike Selection and Nativelike Fluency." In Jack C. Richards and Richard W. Schmidt (eds.), *Language and Communication*. London: Longman, 191–226.

Piccardo, Enrica, Brian North, and Tom Goodier. 2019. "Broadening the Scope of Language Education: Mediation, Plurilingualism, and Collaborative Learning: the CEFR Companion Volume." *Journal of e-Learning and Knowledge Society* 15 (1): 17–36. https://doi.org/10.20368/1971-8829/1612.

Quillbot. 2019. "QuillBot | Free Paraphrasing Tool - Best Article Rewriter." Quillbot.com. 2019. https://quillbot.com/.

Rafajlovičová, Rita. 2012. "The Distribution and Role of Relative Clauses in Different Text Types." *English Matters III (a collection of papers by the Institute of English and American Studies Faculty)*, 11–24.

Rogers, Carmen Villegas, and Frank W. Medley. 1988. "Language with a Purpose: Using Authentic Materials in the Foreign Language Classroom." *Foreign Language Annals* 21 (5): 467–78. https://doi.org/10.1111/j.1944-9720.1988.tb01098.x.

Savenye, Wilhelmina C., Zane Olina, and Mary Niemczyk. 2001. "So You Are Going to Be an Online Writing Instructor: Issues in Designing, Developing, and Delivering an Online Course." *Computers and Composition* 18 (4): 371–85. https://doi.org/10.1016/s8755-4615(01)00069-x.

Savenye, Wilhelmina C. 2005. "Improving Online Courses: What Is Interaction And Why Use It?" *Distance Learning* 2 (6): 22–28.

Schmitt, Norbert, and Bruce Dunham. 1999. "Exploring Native and Non-Native Intuitions of Word Frequency." *Second Language Research* 15 (4): 389–411. http://www.jstor.org/stable/43111242.

Smith, Patricia J., and Tillman J. Ragan. 1999. *Instructional Design.* Upper Saddle River, NJ: Merrill.

Spärck Jones, Karen. 1983. "Compound Noun Interpretation Problems." *Department of Computer Science and Technology, University of Cambridge* 45. https://doi.org/10.48456/tr-45.

Swaffar, Janet K. 1985. "Reading Authentic Texts in a Foreign Language: A Cognitive Model." *The Modern Language Journal* 69 (1): 15–34. https://doi.org/10.1111/j.1540-4781.1985.tb02521.x.

Tai, Hung-Cheng, Mei-Yu Pan, and Bih-O Lee. 2015. "Applying Technological Pedagogical and Content Knowledge (TPACK) Model to Develop an Online English Writing Course for Nursing Students." *Nurse Education Today* 35 (6): 782–88. https://doi.org/10.1016/j.nedt.2015.02.016.

"The IWeb Corpus." n.d. Www.english-Corpora.org. https://www.english-corpora.org/iweb/.

Womack, Anne-Marie. 2017. "Teaching Is Accommodation: Universally Designing Composition Classrooms and Syllabi." *College Composition and Communication* 68 (3): 494–525. http://www.jstor.org/stable/44783578.