

Corpus Integration in L2 Writing Courses: A Cross-Linguistic Didactic Intervention Study

Karla Csürös
West University of Timișoara

Claudiu Gherasim
West University of Timișoara

Mădălina Chitez
West University of Timișoara

Abstract

This study evaluates the effectiveness of corpus-based interventions for enhancing writing skills in English L2 and French L2 among Romanian-speaking students. Following established intervention models, the study involved five stages: initial essay writing, corpus tool training, introduction to target language corpora, essay revision using corpora, and a satisfaction survey. Analysis of linguistic data (e.g., frequency lists, n-grams, and error correction rates) and survey responses from 40 participants reveals improvements in writing accuracy and diversity. Specifically, English L2 students demonstrated enhanced lexical accuracy and varied phraseology, while French L2 students improved syntactic precision and contextual use of academic terms. Both groups showed increased grammatical accuracy, especially in prepositions and articles, through corpus consultation. The findings underscore the pedagogical potential of corpora in writing instruction and the necessity of expanding corpus resources for under-resourced languages like French.

Introduction

As we navigate the current technological revolution, in which AI language models such as those used in ChatGPT become familiar all-purpose tools for both academic and non-academic inquiries, the question arises whether the usefulness of corpora as linguistic supports remains relevant. Until now, corpus-based writing instruction provided the primary means for accessing authentic language use, offering students insights into real-life linguistic patterns. However, with the advent of AI, tools like ChatGPT now offer similar access to language examples, albeit with the added advantage of instant feedback and user interactivity. This shift raises important questions about how these tools might complement or even transform traditional corpus-based approaches.

Previous corpus-based practice has been demonstrated to expand beyond conventional textbook paradigms and offers linguistic as well as pedagogical advantages (Chambers et al. 2011). In language acquisition, the use of corpora refers not only to awareness of corpus methodology but also to actual training regarding the use of concordancing software (Alsaifi, 2022; Cortes, 2011; Chen & Flowerdew, 2018; Fligelstone, 1993; McEnery & Wilson, 1997). This exposure fosters deeper language immersion and facilitates the development of writing skills (Chambers 2005). Additionally, corpus analysis is instrumental in recognizing repetitive patterns within specific contexts, termed as “multicontextual learning” (Cobb, 1997, p. 303), assisting learners in assimilating essential structures for becoming “successful language

learners and language users" (O'Sullivan & Chambers, 2006, p. 52). The synergy between corpus analysis and process-oriented teaching theories empowers learners with the resources and tools they need to construct their understanding (Bernardini, 2002, p. 179), thereby supporting cognitive and metacognitive skills development (O'Sullivan & Chambers, 2006, p. 52). Corpus-based independent exploration of language forms and patterns promotes learner autonomy (Kennedy & Miceli, 2001) which, under the guidance of a teacher, trains self-directed engagement with educational technologies (Benson, 2001, p. 111) and can increase students' motivation (Argyroulis, 2022).

This paper discusses an intervention study on corpus-informed writing instruction in English and French L2 teaching contexts. The study aims to examine the role of corpora in enhancing writing skills in foreign language learning environments. The data was collected from two L2 learner groups with similar L1 backgrounds (i.e., Romanian-speaking undergraduate university students): an L2 English group and an L2 French group. There were 20 students in each group. The intervention study was designed following other successful didactic intervention scenarios (e.g., Bercuci & Chitez, 2019; Chitez & Bercuci, 2020, 2019), consisting of:

- (a) writing an initial paper as an assignment;
- (b) training for corpus tool use;
- (c) introduction to the target language corpora (i.e., English L1 and French L1);
- (d) rewriting the assignment in (a) using corpora in (c); and
- (e) answering a corpus use satisfaction survey.

During the didactic intervention, we introduced students to web-based concordancers for French and English to correct their writing errors or select better options after receiving indirect written feedback from their instructors, discussed further below.

Context

Advantages of specialized corpus consultations compared to AI language models

There are several arguments in favor of accessing specialized corpora as didactic resources. First, such corpora represent authentic linguistic data that has been processed according to clear compilation criteria (Rogobete et al., 2021). They include genre, disciplinary and academic style particularities that can match the user's specific writing needs. In this way, they can be very useful to students. For example, if the user is an economics student and needs to write an academic paper in economics, a corpus containing only such texts would be the perfect instrument to extract field terminology or discipline-specific phraseology. Most importantly, the appropriate linguistic element can be selected from the concordance lists where these elements can be observed in their multiple contexts. By comparison, ChatGPT (or similar tools) can sometimes provide general and non-context-specific examples, with occasional repetition, depending on the linguistic query. Its distinctive writing style, as noted in various tasks (Amirjalili et al., 2024), may influence the development of a student's unique academic voice, raising concerns about preserving individuality in writing (Chitez et al., 2024). However, these shortcomings often depend on the prompt given, as specific and well-crafted prompts can help mitigate issues such as repetition and lack of contextual relevance, improving the overall quality of responses (Dhaini et al., 2023).

Corpus tools also support research-informed teaching practices by allowing educators to analyze common linguistic errors, patterns of vocabulary usage, and areas of textual complexity. These insights help to address learners' specific needs and improve instructional approaches. While AI tools provide rapid and broad feedback, they lack the refined, context-sensitive data needed for detailed analysis and targeted improvements.

In brief, corpus tools and AI tools serve complementary roles in language learning. AI tools excel in providing immediate responses and overcoming writer's block, while corpus tools offer in-depth insights into authentic language patterns, resulting in a more nuanced understanding and application of language.

Corpus consultation: Main functionalities

The concordancer stands as a principal functionality for analyzing corpora. According to Johns (1988, p. 9) a concordancer is defined as a corpus-specific analysis tool which scans extensive textual data to pinpoint a particular element, such as a morpheme, a word, or a phrase, extracting and displaying all instances found alongside their respective contexts of use. The analyzed element is centrally placed in each line, with additional space on both sides, encompassing the adjacent context to facilitate examination of its immediate 'textual environment'. This configuration offers clarity, facilitating the detection of patterns with efficiency (O'Sullivan & Chambers, 2006, p. 50). According to McEnery and Wilson (1997, p. 12), the application of concordances in educational settings can be traced back to 1969 in English for Specific Purposes courses (Aston University, Birmingham). Since the 1980s, studies have highlighted the potential of concordances (O'Sullivan & Chambers, 2006, p. 50) and exploring the impact of corpus consultations on language learning processes or outcomes (Bernardini, 2000, 2002; Cobb, 1997; Johns 1986; Kennedy & Miceli, 2001; Leech & Candlin 1986; Stevens, 1991; Sun, 2003; Tribble & Jones, 1990; Turnbull & Burston 1998). More recently, there has been an increasing focus on integrating corpus consultation in L2 classes to enhance writing skills, leading researchers to explore learners' engagement with corpora in L2 writing and error correction (Gaskell & Cobb, 2004; O'Sullivan & Chambers, 2006; Yoon & Hirvela, 2004), with an emphasis on the importance of developing pedagogically relevant corpora (Braun, 2005).

Error feedback

Recent studies on second-language acquisition have particularly focused on L2 writing, specifically grammar acquisition and the use of concordances for error correction. A growing body of evidence suggests that effective grammar correction in L2 writing classes leads to positive outcomes, as students who consistently apply this method show improvements in accuracy (Chambers et al., 2011; Chandler, 2003; Ferris, 1995, 1997; Ferris & Roberts, 2001; O'Sullivan, 2010).

Error correction may take the form of written corrective feedback, which can be provided through direct or indirect means (Ellis, 2009). Direct feedback involves teachers correcting errors in a traditional way; in contrast, indirect feedback entails marking errors in a sentence using circling, underlining, highlighting, or other visual cues, sometimes accompanied by a verbal reminder or error code, and asking students to make the corrections themselves (Ferris, 2002, p. 63). Furthermore, Ferris and Roberts (2001) emphasize that second-language acquisition research has demonstrated a preference for indirect feedback over direct feedback, as it encourages student engagement and reflection and stimulates their cognitive involvement in the correction process, thereby fostering long-term acquisition. However, other researchers (Bitchener & Knoch, 2009) have presented contrasting results, indicating that direct feedback is more effective in the learning process. The latter argue that indirect feedback often leaves learners feeling frustrated as they do not have the opportunity to rectify their errors. As demonstrated by Crosthwaite et al. (2020), there are no definitive results to establish which method performs better, despite extensive research and numerous meta-analyses comparing the effects of direct and indirect feedback. In this study, we gave students indirect feedback on their writing errors to allow them to discover the correct answers on their own (see fig. 2 for examples).

Methodology

We created a three-stage didactic intervention design to improve first-year students' writing skills, in accordance with other studies of this type (Bercuci & Chitez, 2019; Chitez & Bercuci, 2020, 2019), as follows: (a) writing stage; (b) training session; and (c) rewriting stage. With this design, we were able to assess the students' writing ability and address their explicit needs during the training session. In the first stage (a), the students were given the task of writing a short opinion essay (200–300 words) as a response to a given prompt (an excerpt from the French newspaper *La Voix du Nord*). The L2 French students received the original excerpt, whereas the L2 English students received a translated copy (see fig. 1 below). The topic of the excerpt and of the students' opinion essays was "Is the role of university to prepare for a

profession?”. This topic was chosen for its relevance to the experience of first-year students, who are at the beginning of their academic careers. Both groups of students were given four weeks to complete the assignment at home.

<p>Lisez le texte ci-dessous sur le rôle des universités dans la préparation des étudiants à une profession. Quelle est votre opinion sur le sujet ? Justifiez votre réponse et donnez des exemples pertinents tirés de votre propre expérience. Rédigez un essai d'opinion (entre 200 et 300 mots).</p> <p>Le rôle de l'université est-il de préparer à un métier ?</p> <p>C'est ce que font les universités, de plus en plus. La professionnalisation est entrée dans les établissements universitaires il y a une quinzaine d'années. Avant, ce souci de professionnalisation existait moins, mais il y avait aussi moins d'élèves qui allaient à l'université.</p> <p>Aujourd'hui, on est dans une logique d'insertion professionnelle imposée par la loi et cela se traduit par la création d'un bureau d'insertion (1) dans chaque université. De plus en plus, l'université ne fait que préparer à un métier : elle ne devrait pas se limiter à cela.</p> <p>Bien sûr, on a envie que les étudiants réussissent. Mais l'université n'est pas responsable si les étudiants ne trouvent pas de travail ensuite. On dit : « Ils ont été mal formés ». L'université doit d'abord fournir une culture et un bagage qui vont permettre aux diplômés de s'insérer, mais ce qui se passe après, en termes d'insertion professionnelle, ce n'est plus du domaine de l'université.</p> <p style="text-align: right;">D'après <i>La voix du nord</i></p> <p>(1) Le Bureau d'aide à l'insertion professionnelle recueille et diffuse une offre de stages et d'emplois variée, en lien avec les formations proposées par l'université. Il conseille les étudiants sur leurs problématiques liées à l'emploi et à l'insertion professionnelle.</p>	<p>Read the text below about the role of universities in preparing students for a profession. What is your opinion on the subject? Give reasons for your answer and include any relevant examples from your own experience. Write between 200 and 300 words.</p> <p>Is the role of university to prepare for a profession?</p> <p>This is what universities are doing more and more frequently. Professionalisation has become a point of interest of universities around fifteen years ago. Before, there was less concern about professionalisation, but there were also fewer students going to university.</p> <p>Today, they are in line with the policy of professional insertion imposed by the law, which is translated by the creation of an insertion office¹ in each university. Increasingly, universities are only preparing students for a profession: they should not be limited to that.</p> <p>Of course, the main goal is for students to succeed. But the university is not responsible if students do not find a job afterwards. They say: "They have been badly trained". The university must first of all provide a culture and a background that will enable graduates to find a job, but what happens afterwards in terms of professional integration is no longer the area of the university.</p> <p style="text-align: right;">From <i>The voice of the North</i></p> <p>¹ The Professional Insertion Office collects and distributes a varied offer of internships and jobs, in connection with the courses offered by the university. It advises students on their problems related to employment and professional integration.</p>
--	--

Figure 1. Text prompts for opinion essay writing in French (left) and English (right)

The second stage (b) consisted of a hands-on training session scheduled during a class period of 90 minutes. As many students were unfamiliar with corpus linguistics, they were briefly introduced to the concepts of corpora and concordancers, particularly academic writing corpora. Although there are many web-based concordancers available for English corpora (e.g., *BNCweb*, *BYU-BNC*, *SKELL*), the *LexTutor* concordancers (Cobb, 2024) were chosen for this study due to the online availability of both English and French corpora (*BAWE* for English and *Le Monde* for French), to ensure comparability between students of L2 English and L2 French. Thus, students were taught how to use the *LexTutor* concordancer for English (v.9) and French (v.8.5) to search keywords (with and without wildcards), as well as how to identify associated words (collocates).

Students were then asked to brainstorm methods of using corpora for error corrections, and they successfully anticipated some of the ways in which they would use corpora in their own writing: e.g., correcting grammatical errors, contextualizing phrasal verbs, or finding academic vocabulary. Then, students were given a sample text with errors highlighted and were asked to use the concordancer to correct the mistakes. This meant that students began to familiarize themselves with the given tool, which would be used in the final stage.

The final stage (c) took place after the students received indirect feedback on the first part of their assignment. To be more specific, the instructors highlighted the errors and provided brief comments hinting at what the students should look up in the concordancer. Examples in French and English are included in Figure 2 below. After receiving feedback, students were provided a second class period of 90 minutes and another week in which they could edit their essay after consulting the corpora using *LexTutor*. During the class period, the instructors were available to provide additional feedback if requested, whereas the final at-home work was done independently. Lastly, students were asked to complete a satisfaction survey regarding their experience with corpus consultation methods.

Une question **que** préoccupent la majorité de gens est si l'université **vraiment prépare** les jeunes pour une profession et, bien sûr, si **ce** chose doit être la responsabilité des universités. Il y a beaucoup **des** opinions sur ce sujet, la majorité qui se contrarie, donc nous ne pouvons

However, referring to another quote: "...the university is not responsible if students do not find a job afterwards", I think this one is true, because only **on us depend how** much effort and interest give for what we want and how much we struggle for our dreams.

Check whether the keyword "depend" is associated with "on" to its left or to its right. ¶

Reply Resolve

Figure 2. Samples of indirect feedback for OPINE-FR-1 (mistakes highlighted) and OPINE-EN-1 (mistakes commented)

Survey design

After getting first-hand experience using the web-based concordancers for error correction, at the end of the final session the students were asked to fill out an online survey that noted their linguistic background and measured their experience with corpus use. In a set of six open-ended questions, students were asked to write down their linguistic profile (i.e. degree specialization and year, nationality, native language, level of L2 proficiency). Then, students were asked corpus-related questions, such as “Do you know what a corpus is?”, and “Where did you first encounter the concept of corpus?”. Moreover, they were asked to rate on a scale from 1 (very little) to 4 (very much) how much corpus consultation helped them improve their academic writing (grammar, vocabulary, academic style), and whether they would recommend or want to learn more about academic writing corpora.

Analysis

Because we wanted to verify whether there were any differences between pre- and post-corpus consultation writing, we gathered a total of 80 texts balanced evenly between the two groups as follows: 40 original drafts submitted during the (a) stage and 40 edited final versions written post-corpus consultation during the (c) stage. From these texts, we built a small corpus of student opinion writing, called OPINE, of approximately 21,000 words, with an almost even distribution across pre- and post-consultation for each language, as shown in Figure 3. Note that within OPINE, subcorpus 1 comprises pre-consultation texts, and subcorpus 2 comprises post-consultation texts. Thus in Figure 3 the segment labelled EN-1 (upper left) denotes papers in English at the pre-consultation stage.

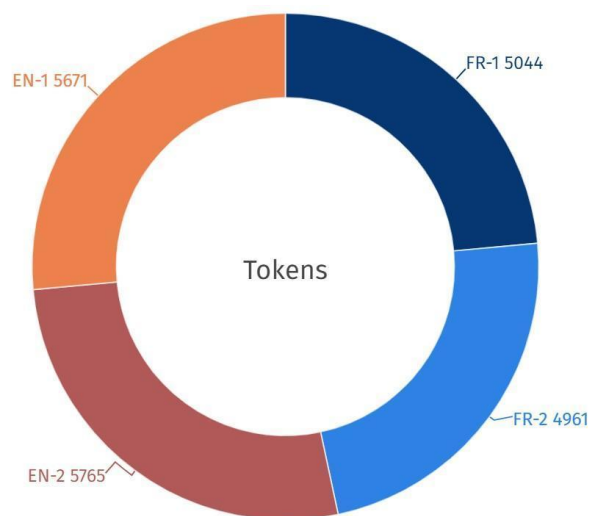


Figure 3. Number and distribution of tokens in OPINE corpus

For the analysis of the OPINE data presented here, *LancsBox* v.6.0 (Brezina et al. 2020) was used to compare the pre- and post-consultation subcorpora for each language. First, we generated frequency lists, that is, the most used words in the texts in ranked order. Then, we looked at changes in the frequency of n-grams between 2 and 5 units. For both lists, we analyzed the use of the most problematic lexico-grammatical items in context using the ‘key word in context’ (KWIC) function.

Frequency lists (FR and EN)

First, we used the frequency lists to compare the pre- and post-corpus consultation subcorpora to check what changes occurred in the students’ use of language (see Tables 1 and 2 below). For both languages, the pre-consultation frequency list revealed **frequent misuse of prepositions, articles, and determiners**:

- In French, common errors included the incorrect use of the prepositions ‘de’ and ‘à’, and the inconsistent application of definite articles, especially the feminine ‘la’ and the

plural 'les'. The French list also showed limited use of linking words. Basic connectors such as 'et' were overused, while more important transition words like 'premièrement' or 'finalement' were never employed.

- In English, 'the' was overused in cases where indefinite articles were more suitable. Furthermore, gendered pronouns ('she', 'he') were misused in contexts referring to inanimate objects, following the Romanian principle of gendering all nouns. When it comes to linking words, students primarily used the same structure ('firstly', 'secondly').

For both languages, the post-consultation frequency list showed improvement in all the areas indicated by the initial frequency list. Correct usage of prepositions increased, with learners making fewer errors in differentiating between 'de' and 'à' in French and 'the' and 'a(n)' in English, thus reflecting a better grasp of grammatical structures. The proper use of articles, determiners, and gendered pronouns was also enhanced, though further improvements are still possible. There was also a small increase in the use and variety of linking words (e.g., introductory words like 'premièrement' in French and conclusory phrases other than 'in conclusion' for English), but nevertheless, learners need to use a wider variety of transition words to indicate an enhanced ability to structure their arguments more coherently.

Thus, it can be said that the increase in the variety and frequency of linking words may underscore the role of corpus consultation in teaching effective discourse markers. Learners also demonstrated a better understanding of the contexts in which each term is appropriately used, reducing previous confusion.

Moreover, in the pre-consultation subcorpora, ***misunderstandings*** were evident in the use of terms related to educational institutions in both languages:

- French learners often used 'université', 'faculté' and 'collège' interchangeably, without recognizing the nuanced differences in their meanings and contexts.
- The same can be said of English learners, who used 'university', 'faculty' and 'college' to refer to the same concept. This can be attributed to the fact that, in Romanian, 'going to university' can be translated using any of the three words, i.e. there is virtually little to no difference between their meanings.

The post-consultation texts displayed a clearer differentiation between the three terms in French and English.

All frequency-based improvements are indicators that the indirect feedback, accompanied by corpus consultation, had given positive results. The clearer distinction between specialized terms demonstrates that enhanced linguistic contextual knowledge, facilitated by guidance in corpus consultation, is beneficial to accurate and context-appropriate language use in L2 writing.

Table 1. Frequency changes in OPINE-FR and OPINE-EN

Token	FR-1 Freq.	FR-2 Freq.	Token	EN-1 Freq.	EN-2 Freq.
de	207	205	a	188	193
et	121	119	of	130	138
la	109	106	university	55	56
les	103	105	it	51	56
à	98	117	this	38	45
université	58	67	college	25	29
collège	4	0	faculty	19	14
premièrement	0	3	secondly	5	2
finalement	0	2	firstly	4	1

N-gram changes (FR and EN)

Natural language includes fixed structures, common word combinations, pre-made expressions, and sentence frames, collectively referred to as 'n-grams' (Baker et al., 2006, p. 122): a sequence of 'n' number of words that repeat often. The exploration of their roles in language development, particularly in L2 writing skills, holds significant importance, as phraseological items serve as crucial indicators of fluent academic writing (Zhang & Li 2021).

Table 2 displays some of the most relevant changes in the use of n-grams from the initial versions of the papers to the final post-consultation submissions.

Table 2. Changes in n-gram frequency in OPINE-FR and OPINE-EN

Use	OPINE-FR-1 vs. OPINE-FR-2	OPINE-EN-1 vs. OPINE-EN-2
Increase ↗	“par exemple”, “la faculté”, “la formation” “à mon avis”, “marché du travail”, “point de vue” “le marché du travail”, “Je suis d'accord avec” “vous permettent de travailler dans”, “vous rend aptes à travailler”	“in the”, “it is”, “do not”, “is not” “the fact that”, “do not have”, “I do not” “on the one hand” “from my point of view”
Decrease ↘	“de la”, “le domaine”, “le rôle”, “un emploi” “je pense que”, “à cause de”, “est très important” “aider ses étudiants de”, “capable de travailler dans un”, “vous rend capable de travailler”	“the university”, “the faculty”, “in conclusion”, “I think” “the university is”, “that the university”, “the role of”, “in my opinion”, “I think that”, “first of all” “in my opinion the”, “on the other hand”, “I think that the” “in conclusion I think that”
Stable →	“du point de vue”, “le rôle de l'université”	“I believe that the”

For French, the frequency of ‘de la’ decreased, indicating a shift towards more precise or varied prepositional phrases. Reduced usage of ‘le domaine’ and ‘le rôle’ suggests a decrease in generic phrases, possibly due to an expanded vocabulary, and an increase in synonym variety or more specific language. The increase in the transition phrase ‘par exemple’ points to improved ability to set up arguments by providing illustrative examples, which is a key aspect of argumentative skills. While this does not necessarily reflect the overall quality of the argument itself, it suggests progress in structuring and supporting ideas within L2 writing. The higher frequency of ‘la faculté’ and ‘la formation’ indicates improved contextual understanding of educational terminology and a better grasp of specific academic and professional contexts. There is also a drop in the phrase ‘je pense que’ which suggests a shift to more varied expressions of opinion. Reduced usage of ‘à cause de’ and ‘est très important’ illustrates that learners are using more varied casual expressions and more nuanced evaluative language. The rise in frequency of the 3-grams ‘à mon avis’, ‘point de vue’ and ‘marché du travail’ shows that learners are adopting diverse expressions for stating opinions, indicating improved argumentative skills and a more precise application of terms related to professional contexts. The decrease in the frequently incorrect structure ‘aider ses étudiants de’ demonstrates a better understanding of the language and the use of more grammatically correct phrases. Consistent usage of structures such as ‘du point de vue’ and ‘le rôle de l'université’ implies consistent application of phrases within academic contexts, which indicates their stability in learners’ lexicon. The increased frequency of ‘le marché du travail’ and ‘Je suis d'accord avec’ reinforce the conclusions reached in the case of the 3-grams above, showing an improved ability to express agreement (and disagreement) in a given L2 context. Both ‘capable de travailler dans un’ and ‘vous rend capable de travailler’ appeared frequently in the pre-consultation texts, indicating learners’ reliance on these specific structures to convey the idea of being capable of working in a particular context or under specific conditions. The decrease in the use of these 5-grams may be attributed to an increase in lexical and syntactical variety in learners’ writing.

For English, a decrease was noted in the use of the definite article ‘the’ with ‘university’ and ‘faculty’, proving that students became more aware of the contextual differences between ‘the’ and ‘a(n)’. The students also removed any contracted verb forms and replaced them with the long forms ‘it is’, ‘(I) do not (have)’ and ‘is not’. The reduction in the overuse of ‘in conclusion’, ‘I think (that the)’, both separate and combined, shows that students were able to assimilate more varied academic writing structures used for argumentation, such as ‘from my point of view’ as opposed to ‘in my opinion the’. The constant use of ‘I believe that the’ may display the students’ perception of it as more academic than ‘I think’. Interestingly, the use of ‘on the other

hand' decreased, whereas the use of 'on the one hand' increased; this is perhaps again due to the Romanian bias for the first form ('pe de altă parte'), as opposed to the initial structure.

Post-corpus consultation, both L2 French and L2 English students encountered and adopted a broader range of expressions to convey similar meanings, leading to a reduction in reliance on repetitive structures.

Error Categories

There was a total of 328 errors in OPINE-1 (i.e., the pre-consultation subcorpus) across the two languages (fig. 4). Of the total of 226 errors in French, 118 are grammatical (52%), 28 are lexical (12%), 15 are syntactic (7%) and 65 are substance related (29%). Students successfully made 221 corrections, resulting in an accuracy rate of 98%. For English, there were a total of 102 errors, of which 47 were successfully corrected (88% accuracy). These were mainly related to grammar (40%), primarily the misuse of prepositions and verb forms, as well as lexicon errors (28%) related to word choice.

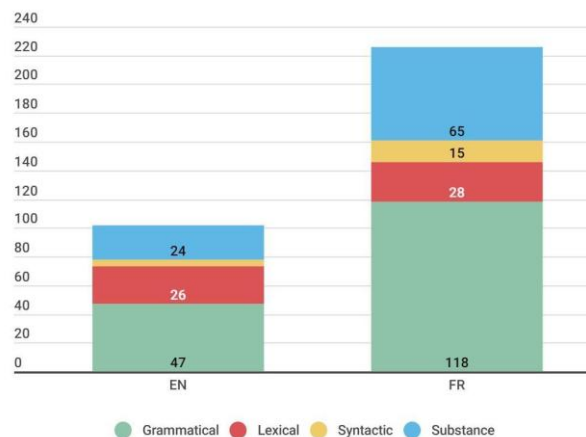


Figure 4. Number of errors in OPINE-1

The largest category of L2 French errors (see fig. 5) indicated that the majority fall under grammatical issues, with specific challenges in prepositions, gender and agreement, and verb forms. This category highlighted difficulties with matching the gender of nouns and adjectives or pronouns and ensuring agreement, which is a common issue in L2 gendered grammar. Another area of grammatical errors was related to prepositions, especially in the context of prepositional verbs that require the prepositions 'à' (e.g., 'penser à') or 'de' (e.g., 's'occuper de'). Other grammatical errors indicated challenges with conjugating verbs correctly based on tense, aspect, or mood, problems with forming plural nouns or ensuring agreement between subjects and verbs, and incorrect forms of pronouns, definite and indefinite articles, and adverbs. Lexical and syntactic errors in L2 French showed difficulty in choosing appropriate words and structuring sentences correctly, which affects the meaning, clarity, and comprehensibility of their writing. Substance errors in L2 French, the second largest category, consisted of frequent misspellings and errors in using correct accents, which are decisive in French as they can change the meaning of a word and sentence (e.g., 'du' and 'dû'; 'ou' and 'où'; 'la' and 'là'; 'cote', 'côte' and 'coté').

Similarly, in L2 English, the largest number of errors were grammatical (see fig. 5). Most were related to the use of incorrect verb forms (i.e., tenses and number), the misuse of pronouns (particularly using gendered pronouns for genderless nouns), as well as improper use of articles (difficulty in distinguishing between the contextual need for 'a(n)' versus 'the'). There were also many lexical issues, especially when it comes to false friends (e.g., the Romanian 'actual' meaning 'current' and the English 'actual' meaning 'real') or phrases being translated literally (e.g., the Romanian phrase 'un an de pauză', meaning 'gap year', being translated as 'a year

of break'). Students also frequently misspelled words, often integrating the Romanian spelling (e.g., 'example' being written as 'exemple' from the Romanian 'exemplu').

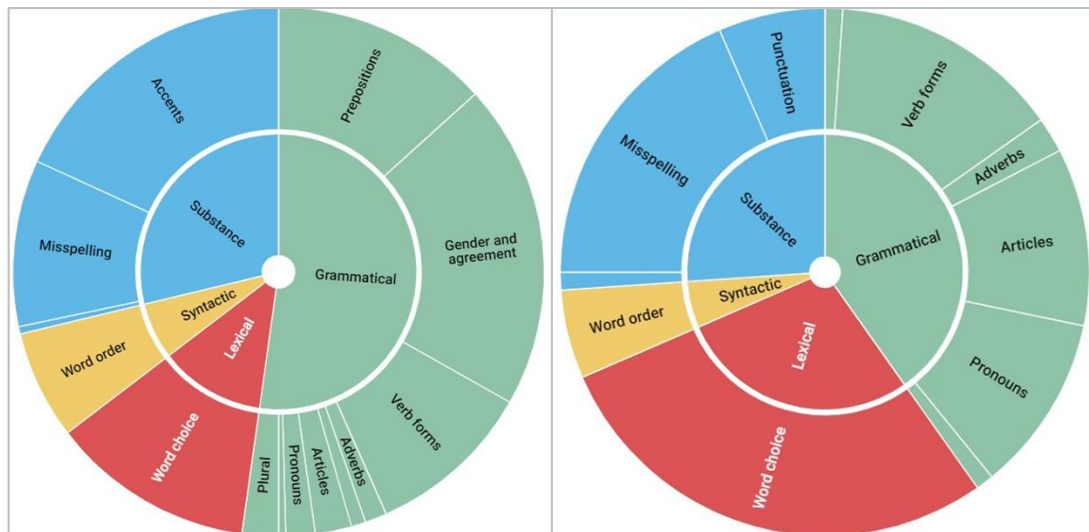


Figure 5. Error categories in OPINE-FR-1 (left) and OPINE-EN-1 (right)

Analyzing and quantifying errors in the post-consultation French corpus (OPINE-FR-2), we noticed that students enhanced the overall grammatical correctness and text coherence (table 3). They also ensured proper verb usage, especially past tense constructions, leading to increased coherence and cohesion between sentences. Rearranging the word order improved syntactic accuracy, making the sentences more natural and comprehensible. Corrections to partitive expressions conveyed the intended meaning more accurately, while changes related to idiomatic expressions and fluency ensured smoother language usage. Corrected proper names that align with standard naming conventions managed to make the reference more accurate and recognizable. Additionally, improvements in gender and number agreement, pronoun placement, and preposition usage enhanced grammatical accuracy, clarity, and readability.

Table 3. Examples of changes in OPINE-FR

OPINE-FR-1	OPINE-FR-2
être une étudiant de lettre je pense	étant une étudiante de lettres, je pense
Nous sommes faits des	Nous avons fait des
d'accord avec ce que dit le texte	d'accord avec ce que le texte dit
Ils gagnent l'expérience	Ils gagnent de l'expérience
dans mon avis	à mon avis
en aider (...) de mieux s'intégrer	en aidant (...) à mieux s'intégrer
sa rôle dans la vie du adolescents	son rôle dans la vie des adolescents
l'Université occidentale de Timișoara	l'Université de l'Ouest de Timișoara
à cause de la manque du temps	à cause du manque de temps
pour aider nous	pour nous aider

For L2 English (table 4), taking a closer look at the textual characteristics in the samples of corrected errors, we observed specific improvements in students' writing. The revised texts exhibited improved subject-verb agreement and the inclusion of subjects in sentences, addressing errors stemming from the optional nature of subjects in Romanian syntax (e.g., 'will guarantee' revised to 'it will guarantee'). Additionally, the texts demonstrated enhanced lexical choices and a reduction in L1 transfer, particularly in fixed phrases; for example, the Romanian phrase 'a lua un examen' (literally 'take an exam') was revised to the more appropriate English equivalent 'pass an exam.' There were also notable improvements in article usage, as seen in revisions like 'is Faculty of Medicine' to 'is the Faculty of Medicine', and word order, as in 'how actually education should work' revised to 'how education should actually work'. These

observations reflect measurable changes in linguistic accuracy while consulting a corpus even if the dataset consists of one-off tasks of 200–300 words.

Table 4. Examples of changes in OPINE-EN

OPINE-EN-1	OPINE-EN-2
Moreover, will guarantee that	Moreover, it will guarantee that
after finishing the faculty I will	after finishing college, I will
is Faculty of Medicine	is the Faculty of Medicine
how actually education should work	how education should actually work
if I did not take it afterwards	if I did not pass it afterwards
I think that the university prepares	I think that university prepares
we start to know	we get to know
learn also to put them in practice	also learn to put them in practice
me and my colleagues learned a lot	my colleagues and I learned a lot
to make us to realize	to make us realize

From the above-mentioned error categories that were most frequently made by L2 French and L2 English students, alongside the positive changes that they made in their final submissions (i.e., post-corpus consultation texts), we can see that corpus-consultation methods strongly impacted their writing. Most common error types can be attributed to the students employing calques (i.e., word-for-word translations from L1 Romanian to their L2 language). Direct contact with target language corpora allows students to understand contextual differences between words and phrases in their L1 and L2. Even though this deeper understanding is noticeable from our perspective as instructors, we also wanted to understand the students' experience with corpus use. That is why we also asked them to complete a satisfaction survey, described below.

Survey Results

The final part of our study encompassed an online survey meant to gauge the students' satisfaction and overall experience using corpora for writing. The survey was shared after the students received feedback on the final drafts of their papers. All 40 students who took part in the intervention study also participated in the survey. Out of the 40 participants, 38 identified as ethnically Romanian, and 2 as ethnically Moldovan. Moreover, 39 participants had Romanian as a mother tongue, with 1 claiming Slovakian as their native language. Thirty-six participants were first-year students, with four French learners being in their second year and retaking the class. Therefore, we can say that this study focused on native Romanian learners' experiences with L2 English and L2 French in the early stages of their undergraduate careers.

Both groups of students reported a lack of prior knowledge when it comes to corpora, yet they were confident that they knew the meaning of the word after the in-class training session. They also wanted to learn more about corpus linguistics, showing that they found this intervention useful. More explicitly, when answering question Q3, 'How have corpus-consultation methods improved your writing?', both L2 French and L2 English students were very positive about their experiences. As per Figure 6, students found corpus consultation to be very helpful when it comes to grammar and single words, as well as expressions and academic style.

In the final part of the survey, students were requested to offer their opinions on corpus consultation methods in an open-ended question. Both the L2 English and L2 French students were quite positive in their reflections, highlighting some key benefits of using academic corpora for writing, both cognitive and practical. First, students noted the reliability of academic corpora as 'a trustworthy academic support' and a 'safe source of information'. Then, they appreciated an increase in learning independence, noting that they no longer fully depended on the instructor to give them feedback. Instead, corpora 'helped [them] in seeing and correcting [their] own grammatical and orthographic mistakes' and gave them the opportunity to 'revise [their] mistakes on [their] own'. Moreover, students also noted the lexical, grammatical, and syntactic improvements that using academic corpora brought them. For instance, students claimed to

have ‘improved vocabulary’, ‘better and more varied vocabulary’, as well as exposure to ‘other phrases than the classic ones’. Contextual information provided by the KWIC function proved to be particularly useful for students, who remarked that they learned to ‘us[e] words in the right contexts’. In this case, the role of academic corpora was dual: students discovered ‘more contexts for certain words’, but also ‘found the proper words for a given context’.

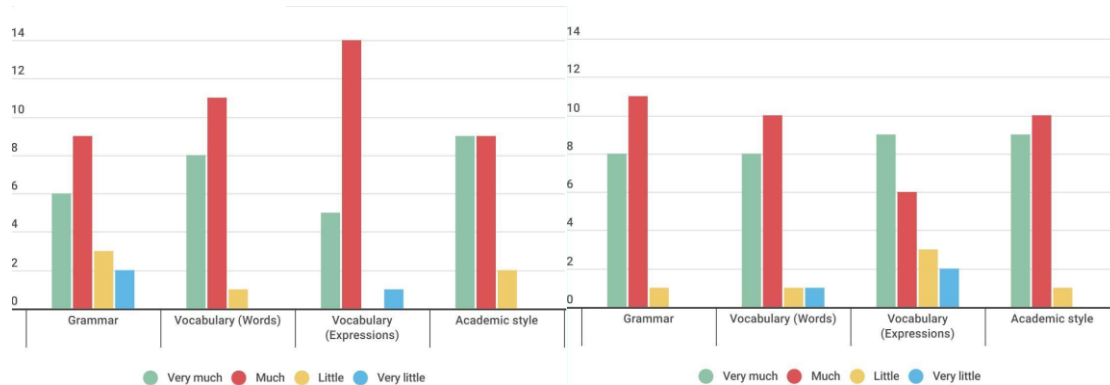


Figure 6. Responses to Q3 [How have corpus-consultation methods improved your writing?] for French (left) and English (right)

Conclusions

Both linguistic data and survey data gathered from 40 informants emphasize the need to adapt corpus-based approaches in writing classes to the specifics of the language. It has been found that both L2 English and L2 French students can use corpora for lexical and grammatical error correction. Additionally, it appears that L2 English students mainly were able to improve their academic phraseology use through corpora consultation, whereas L2 French learners were able to improve in matters related to syntax. Students generally found corpus consultation to be a useful approach, and were willing to learn more about using corpora in their writing. It should be noted that while English is a digitally well-resourced language, using corpora for writing in French (Chambers & O’Sullivan, 2004) is less common, since fewer corpora are available. For this reason, even if the comparison between teaching L2 English and L2 French with the help of corpora is not balanced due to access to different resources, the results of the study can also be used as indicators of the necessity to build more specialized or general corpora in various languages, such as French.

References

- Alsahafi, M. (2022). Corpora and concordancing in second language vocabulary learning. *Theory and Practice in Language Studies*, 12(12), 2651–2657. <https://doi.org/10.17507/tpls.1212.23>
- Amirjalili, F., Neysani, M., & Nikbakht, A. (2024). Exploring the boundaries of authorship: A comparative analysis of AI-generated text and human academic writing in English literature. *Frontiers in Education*, 9. <https://doi.org/10.3389/feduc.2024.1347421>
- Argyroulis, V. (2022). Investigating student motivation in the use of corpus concordancing in ESP learning at university level. *ESP Today*, 10(1), 71–98. <https://doi.org/10.18485/esptoday.2022.10.1.4>
- Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh University Press.
- Benson, P. (2001). *Teaching and researching autonomy in language learning*. Longman.
- Bercuci, L., & Chitez, M. (2019). A corpus analysis of argumentative structures in ESP writing. *International Online Journal of Education and Teaching*, 6(4), 733–747. <https://www.iojet.org/index.php/IOJET/article/view/655/431>
- Bernardini, S. (2000). Systematising serendipity: Proposals for concordancing large corpora with language learners. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 225–234). Peter Lang.
- Bernardini, S. (2002). Exploring new directions for discovery learning. In B. Kettemann & G. Marko (Eds.), *Teaching and learning by doing corpus analysis* (pp. 165–182). Rodopi.
- Bitchener, J., & Knoch, U. (2009). The relative effectiveness of different types of direct written corrective feedback. *System*, 37(2), 322–329. <https://doi.org/10.1016/j.system.2008.12.006>
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47–64. <https://doi.org/10.1017/S0958344005000510>
- Brezina, V., Weill-Tessier, P., & McEnery, T. (2020). #LancsBox (Version 6.0) [Computer software]. <http://corpora.lancs.ac.uk/lancsbox/> [Editor's note: a new version of LancsBox was released last year along with a new URL to which the old URL redirects – <https://lancsbox.lancs.ac.uk/>.]
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology*, 9(2), 111–125. <https://doi.org/10.125/44022>
- Chambers, A., Farr, F., & O'Riordan, S. (2011). Language teachers with corpora in mind: From starting steps to walking tall. *The Language Learning Journal*, 39(1), 85–104. <https://doi.org/10.1080/09571736.2010.520728>
- Chambers, A., & O'Sullivan, Í. (2004). Corpus consultation and advanced learners' writing skills in French. *ReCALL*, 16(1), 158–172. <https://doi.org/10.1017/S0958344004001211>
- Chen, M., & Flowerdew, J. (2018). A critical review of research and practice in data-driven learning (DDL) in the academic writing classroom. *International Journal of Corpus Linguistics*, 23(3), 335–369. <https://doi.org/10.1075/ijcl.16130.che>
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12(3), 267–296. [https://doi.org/10.1016/S1060-3743\(03\)00038-9](https://doi.org/10.1016/S1060-3743(03)00038-9)

- Chitez, M., & Bercuci, L. (2019). Data-driven learning in ESP university settings in Romania: Multiple corpus consultation approaches for academic writing support. In F. Meunier, J. Van de Vyver, & L. Bradley (Eds.), *CALL and complexity: Short papers from EUROCALL 2019* (pp. 75–81). Research-publishing.net. <https://doi.org/10.14705/rpnet.2019.38.989>
- Chitez, M., & Bercuci, L. (2020). Calibrating digital method integration into ESP courses according to disciplinary settings. *New Trends and Issues Proceedings on Humanities and Social Sciences*, 7(1), 20–29. <https://doi.org/10.18844/prosoc.v7i1.4862>
- Chitez, M., Frîncu, S. I., & Rogobete, R., & Terian, S.-M. (2024, 24–25 October). *A strong case for corpus-related digital humanities: Academic writing corpora as intangible cultural heritage assets* [Paper presentation]. ELABCHROM Conference 2024, Sibiu, Romanian. <https://grants.ulbsibiu.ro/elabchrom/conference-2024/>
- Cobb, T. (2024). *Concordancers*. Compleat Lexical Tutor. <https://www.lextutor.ca/conc/>
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3), 301–315. [https://doi.org/10.1016/S0346-251X\(97\)00024-9](https://doi.org/10.1016/S0346-251X(97)00024-9)
- Cortes, V. (2011). Genre analysis in the academic writing class: With or without corpora? *Quaderns de Filologia: Estudis Lingüístics*, 16, 65–80. <https://turia.uv.es/index.php/qfilologia/article/view/3932>
- Crosthwaite, P., Storch, N., & Schweinberger, M. (2020). Less is more? The impact of written corrective feedback on corpus-assisted L2 error resolution. *Journal of Second Language Writing*, 49. <https://doi.org/10.1016/j.jslw.2020.100729>
- Dhaini, M., Poelman, W., & Erdogan, E. (2023, 4–6 September). Detecting ChatGPT: A survey of the state of detecting ChatGPT-generated text [Paper presentation]. *Proceedings of the 8th Student Research Workshop associated with the International Conference Recent Advances in Natural Language Processing*. INCOMA Ltd. <https://aclanthology.org/2023.ranlp-stud.1/>
- Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107. <https://doi.org/10.1093/elt/ccn023>
- Ferris, D. (1995). Can advanced ESL students become effective self-editors? *CATESOL Journal*, 8(1), 41–62. <https://escholarship.org/uc/item/6k52f2d3>
- Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, 31(2), 315–339. <https://doi.org/10.2307/3588049>
- Ferris, D. (2002). *Treatment of error in second language student writing*. University of Michigan Press.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10(3), 161–184. [https://doi.org/10.1016/S1060-3743\(01\)00039-X](https://doi.org/10.1016/S1060-3743(01)00039-X)
- Fligelstone, S. (1993). Some reflections on the question of teaching, from a corpus linguistics perspective. *ICAME Journal*, 17, 97–109. https://icame.info/icame_static/archives/No_17_ICAME_Journal_index.pdf
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3), 301–319. <https://doi.org/10.1016/j.system.2004.04.001>
- Johns, T. (1986). Micro-concord: A language learner's research tool. *System*, 14(2), 151–162. [https://doi.org/10.1016/0346-251X\(86\)90004-7](https://doi.org/10.1016/0346-251X(86)90004-7)

- Johns, T. (1988). Whence and whither classroom concordancing? In T. Bongaerts, P. de Haan, & S. Lobbes (Eds.), *Computer applications in language learning* (pp. 9–34). Foris Publications. <https://doi.org/10.1515/9783110884876-003>
- Kennedy, C., & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning & Technology*, 5(3), 77–90. <https://doi.org/10.125/44567>
- Leech, G., & Candlin, C. (1986). *Computers in English language teaching and research*. Longman.
- McEnergy, T., & Wilson, A. (1997). Teaching and language corpora (TALC). *ReCALL*, 9(1), 5–14. <https://doi.org/10.1017/S0958344000004572>
- O'Sullivan, Í. (2010). Using corpora to enhance learners' academic writing skills in French. *Revue française de linguistique appliquée*, 15(2), 21–35. <https://shs.cairn.info/journal-revue-francaise-de-linguistique-appliquee-2010-2-page-21>
- O'Sullivan, Í., & Chambers, A. (2006). Learners' writing skills in French: Corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15(1), 49–68. <https://doi.org/10.1016/j.jslw.2006.01.002>
- Rogobete, R., Chitez, M., Mureşan, V., Damian, B., Duciuc, A., Gherasim, C., & Bucur, A.-M. (2021). Challenges in compiling expert corpora for academic writing support. *Conference Proceedings – 11th International Conference The Future of Education* (pp. 409–414). Filodiritto Editore. <https://conference.pixel-online.net/FOE/files/foe/ed0011/FP/6917-ICT5260-FP-FOE11.pdf>
- Stevens, V. (1991). Concordance-based vocabulary exercises: A viable alternative to gap-fillers. *English Language Research Journal*, 4, 47–63. https://vancestevens.com/papers/1991_johns_king/vocabex91.htm
- Sun, Y.-C. (2003). Learning processes, strategies, and web-based concordancers: A case study. *British Journal of Educational Technology*, 34(5), 601–613. <https://doi.org/10.1046/j.0007-1013.2003.00353.x>
- Tribble, C., & Jones, G. (1990). *Concordances in the classroom: A resource book for teachers*. Longman.
- Turnbull, J., & Burston, J. (1998). Towards independent concordance work for students: Lessons from a case study. *On-CALL*, 12(2), 10–21.
- Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, 13(4), 257–283. <https://doi.org/10.1016/j.jslw.2004.06.002>
- Zhang, X., & Li, W. (2021). Effects of n-grams on the rated L2 writing quality of expository essays: A conceptual replication and extension. *System*, 97. <https://doi.org/10.1016/j.system.2020.102437>