研究ノート

# The Use of Concordances in EFL/ESL Writing Instruction.

**Peter PARISE**

## ABSTRACT

The Support Writing Course of the Kanagawa Institute of Language and Culture offers an opportunity to practice academic writing for Japanese teachers of English in junior high and high school in Kanagawa. Due to the unique nature of the course which is conducted mostly on a Moodle website, this allows a greater range of feedback. One type of feedback used in the course is a data-driven approach, which makes use of concordance data for the learners to reflect on their errors and self-edit. Looking at literature reviews and studies involving error correction and data-driven learning, future methodologies and research regarding this approach for the course will be considered.

## Introduction

Of the many courses offered at the Kanagawa Prefectural Institute of Language and Culture's In-Service Teacher Training Division, there is one course offered which serves to help Japanese teachers of English with one often neglected skill: academic writing.

Due to a lack of knowledge of the essay format, and a lack of confidence in their own skill as an English writer, teachers might avoid adopting writing activities in their classes and the cost of this avoidance eventually impacts the students. The Support Writing Course (英文ライティング添削講座) was created for English teachers in junior and senior high schools with the hope that by giving opportunities to practice writing, this much needed confidence and knowledge would be transferred to the classroom.

Another vital role that this course fulfills is the very fact that the participants are exposed to a web-based environment via a secured Moodle website This format allows for more flexibility for instruction than would be possible in a conventional classroom due to the availability of Internet resources for supporting writing. The value of this exposure cannot be overlooked. One such opportunity is the ability to expand feedback to allow students to reflect on their errors in writing.

Our approach to this course is to provide a variety of feedback for the participants. When we receive an assignment, we read and mark it for errors; this feedback approach is called *direct corrective feedback,* which entails marking with a red pen on paper. The error is crossed out and the correct form is given. The same kind of procedure works by using the "track changes" (変更履歴の記録)feature in Microsoft Word. In order to supplement

this feedback, we also offer audio conferencing by recording a personalized message on the website so the participants can listen to explanations from the instructor as to why they received certain corrections in their text. The audio message serves to give greater meta-cognitive feedback, as well as moral support.

Most teachers when teaching writing find that they must give some form feedback to students, most of the time as direct corrective feedback in the form of red pen marks on paper as described above. This can be a labor intensive process with dubious benefits. This paper suggests an approach to feedback which allows the learner to learn from their errors. This process is described as *data-driven learning,* which allows students to examine samples of language in an inductive rather than deductive process.

For the Support Writing Course this third type of feedback was chosen to encourage the participants to be more engaged in the self-editing process. The main question of this paper is: Are there definitive benefits to this feedback based on previous studies involving concordances in writing instruction? To answer this question two sets of literature will be examined: studies of EFL writing regarding error correction and literature on data-driven learning in the language classroom. After establishing this backdrop, this paper will reflect on research possibilities for our writing program and beyond.

**Writing instruction: positions on error correction.**

The position error has held in EFL writing instruction has gone through shifts of focus. The major question in the discipline is if errors should be corrected at all? The debates between Truscott (1996) and Ferris (1999) regarding the utility of error feedback exemplify this tension. Truscott cited numerous studies which compared the differences between groups of students in writing courses that received grammatical correction versus groups without correction, noting that there were no significant differences discovered in those studies. He concluded that grammar correction has "has no place in writing courses and should be abandoned" (p. 328) and that the focus for students both for L1 and L2 writing should be exclusively on the content itself. Truscott claimed that error correction could be harmful in some cases such as the observations in Sheppard (1992), in which the students avoided a structure due to the instructor's focus on grammar. Ferris (1999) in her response Truscott, called attention to the fact that his comparison had limitations: for example the studies he had compared had different types of participants, both L1 and L2 learners, which could not lead to generalizability. A point that she did agree with is his position that different approaches are necessary for feedback on errors and gave some guidelines with emphasis by the author.

….that students can be successfully taught to *self-edit* their own texts if they are (a) focused on the importance of editing; (b) *trained to identify and correct patterns of frequent and serious errors;* and (c) given explicit teaching as needed about the rules governing these patterns of errors. I have also argued *that indirect error correction* (identification of errors) is *preferable to direct correction* (teacher correction of student errors). (Ferris, 1999, p.5)

In short, the treatment of error in writing needs to go beyond just one type of feedback. There needs to be a variety of approaches that the teacher must utilize. Of note is her attention to *indirect* error correction since the direct correction approach may be quite limited in helping learners. One such study was Robb, Ross, & Shortreed (1986), in which they compared four types of feedback for the writing of Japanese students of English. They found that more direct methods did "not tend to produce results commensurate with the amount of effort required of the instructor to draw the student's attention to surface errors." (p. 88).

In more contemporary studies, the issue of the effectiveness of direct correction is still in question. One such study is Van Beuningen, De Jong, & Kuiken (2012). Their investigation into the effectiveness of direct and indirect feedback found that both kinds are effective for different aspects of writing. Direct feedback was helpful for grammatical accuracy whereas indirect feedback was helpful for non-grammatical aspects of writing. Shintani and Ellis (2013) in contrast, found that when compared to an alternative form of writing assessment, direct corrective feedback improved neither implicit nor explicit knowledge necessary for writing.

**Data-driven learning**

An approach that can contrast direct corrective feedback which can facilitate editing, is indirect and makes use of examples to allow the student to identify their errors fits a description of *data-driven learning* (DDL). The definitive article about this approach is Johns (1991) and is known for the oft quoted statement that "research is way too important to leave to the researcher." (p. 2) which means that learners should be allowed to "discover" language rather than receive explanations from the teacher or in the case of the writing class, receive corrections. The learner develops according to Johns, "strategies for discovery" in order to "learn how to learn". Johns describes the role of the computer in this approach is not like that of a teacher but instead that of an "informant" which serves to guide, provided the learner can ask the right questions.(ibid., p. 1)

Data-Driven learning is as an inductive approach for learning about lexical or grammatical forms where the student takes the role of investigator and though examples, determines the rules of language. The main source of these examples comes from an electronic collection of texts called a corpus. The material which builds a corpus is culled from authentic contexts where language is used naturalistic settings. The output for the examples which come from searching corpora is called a concordance (Figure 1) or "keyword in context" (McEnery & Hardie, 2012, p. 35) in which the *node*, the word under investigation, is lined up with other instances of the word to allow analysis. The researcher can look at this concordance and read from the top, bottom, and the left and right sides of the node to find patterns of how the node word is used in the language comprising the corpus.
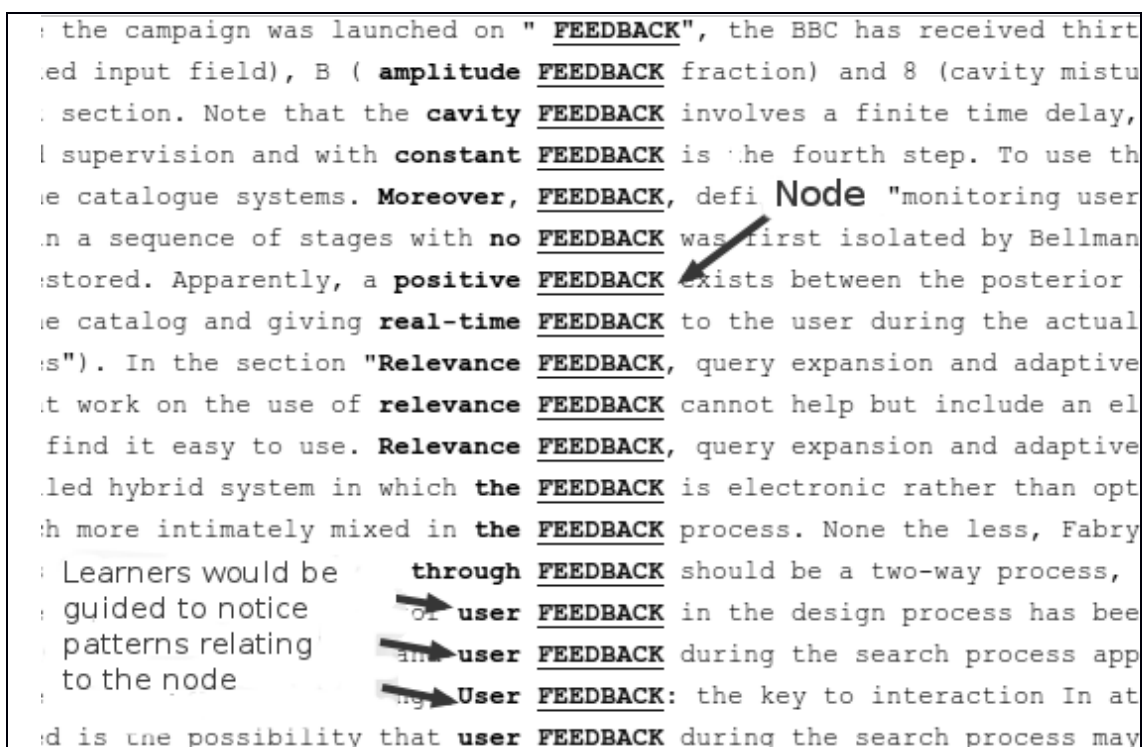


Figure 1. Sample concordance from the Compleat Lexical Tutor.

Since a detailed explanation of corpora and concordances goes beyond the scope of this paper, I recommend reading O'Keeffe, McCarthy, & Carter (2007) for a teacher accessible explanation and McEnery, Xiao, & Tono (2006) or McEnery & Hardie (2012) for a more advanced treatment of the subject.

Data-driven learning can be encapsulated by two approaches: one of the most common as mentioned in the literature review by Chambers (2007, p. 4), is that a majority of researchers have encouraged their participants to directly consult corpora either as an

installed application on the computer or on an Internet based corpus via a web browser. This entails that the learners independently search a corpus themselves in the process of writing. When they encounter a problem related to form or word choice, they consult the corpus as an addition to conventional sources such as dictionaries and thesauruses. In some studies the learners consult large, publicly available corpora available on the Internet such as the British National Corpus which contains a wide variety of genres. Other studies utilize a corpus built by the teacher based on the learners' needs. These specialized corpora can be fine tuned to aid students of specific types of professions such as business writing, engineering, and the legal field (Flowerdew, 2010, pp. 448-449).

The other approach is a mediated use of the data from a corpus. The learners do not directly consult corpora but interact with materials produced from the teacher using the information obtained. Flowerdew describes this as *pedagogic processing* as advocated by Widdowson (1991) regarding the issue of transferability of corpus data to language learning materials. (Flowerdew, 2009, pp. 403–405) This mediation can be in the form of paper based materials (Boulton, 2010) or marking written text to lend a hint of what a student should check in a corpus (Gilmore, 2009), or using hyperlinks that offer ready-made concordances for the learner to use (Gaskell & Cobb, 2004).

Early surveys of DDL (Chambers, 2007; Boulton, 2008) take into account the methods used to research the approach and their effects. Chambers (2007) reviewed 12 studies and categorized them by either qualitative or quantitative methods. She noted that a majority of the studies focused on qualitative data (student reactions, opinions of using concordances, observations of the researcher) and only a few which were quantitative, of note Gaskell and Cobb (2004) and Yoon and Hirvela (2004), the focus in these reports were on the positive responses by their students and how the students saw concordances as a viable tool for their writing (pp. 7-8). Boulton (2008) reviewed 50 studies on DDL by a similar design as Chamber's and made a similar comparison: qualitative reports, especially those involving student reactions were the majority which focused on "annex questions such as what learners do or whether they like doing it, or how effective corpora can be as a reference tool in writing, translating or error-correction rather than as a learning tool." Those that reported quantitative data posted positive results but were unfortunately too slight to make a generalization (p. 85).

The qualitative and quantitative studies above show that students respond positively to the approach. Chambers (2007) in her literature review noted two sets of learner responses to using corpus data in terms of positive response. She divided these into two broad categories. One relating to their language learning experience: the students responded well to the data they obtained and felt that it was more authentic and relevant to

their needs. They also liked the fact that there was an abundance of examples to work from and considered this to be a major advantage over dictionary sources. The other category was in terms of affective elements: learners enjoyed the self-directed nature of the approach, the autonomy and also a felt it was a quick way to check when it came to word or grammatical usage.

These results are tempered with negative responses as well. In the same volume Chambers noted that there were tendencies for learners to be confused with words/phrases due to context. Working with corpora was found to be difficult in that it took time to find satisfactory results that a learner searches for, and that process can be tedious. Also the learners have found "very common and very rare words frustrating" because the output might be above or below their linguistic level. Another issue was trainabilty: the process of getting students accustomed to using the tools can be a challenge (Chambers, 2007, pp. 11–12).

Recent studies involving concordances and writing that have come after these reviews include Abu Alsharr and AbuSeileek (2013), in which they compared three groups of learners: a group that used concordances, another group which was trained to use the grammar and spell checker in Microsoft Word, and a control group that used neither. The results of their study yielded significant findings in particular the concordance group while not as accurate compared to the word processing group in grammar, spelling, and synonym use, this group was better in terms of writing, word collocations, word connotations, and phraseology (pp. 69-70).

Mull (2013), performed a longitudinal study of four learners who accessed and used concordances to peer edit their essays. Using screencasts along with other techniques such as audio recording she documents how her learners interact with this approach (p. 49). Her findings show a highly detailed account of how a student responds to a concordance and corrects based on their own theories about the language (p. 50).

**Concordance feedback the Support Writing Course**

The process of offering data-driven feedback to the participants' writing errors is based on the approach described in Gaskell & Cobb (2004), in which mediated concordances were used as a way to train learners to conduct direct concordance searches on their own. Their approach was to highlight in a document the section of the error and provide, initially on paper and then with hyperlinks in a Word file, concordances with the correct form given in multiple examples. With this the learner inducts the correct form, makes the correction and returns it to the instructor in the form of a feedback loop.

The procedure of our Support Writing Course entails finding an error which may lend itself to concordance feedback, such as a misused vocabulary item, morphology, or even syntax. The instructors used either the concordance feature of the Compleat Lexical Tutor (Cobb, 2013) which has a hyperlinking feature already offered in the concordance section of the website itself. In addition it provides access to large web-based corpora such as the British National Corpus, and the Brown Corpus plus corpora based on graded readers. In addition the Corpus of Contemporary American English (Davies, 2010) was also consulted to obtain sets of examples.

The area in question is highlighted, and using the "comment feature" in Microsoft Word, a comment is added asking the participant to rewrite the highlighted section based on examples found in the link provided (Figure 2)
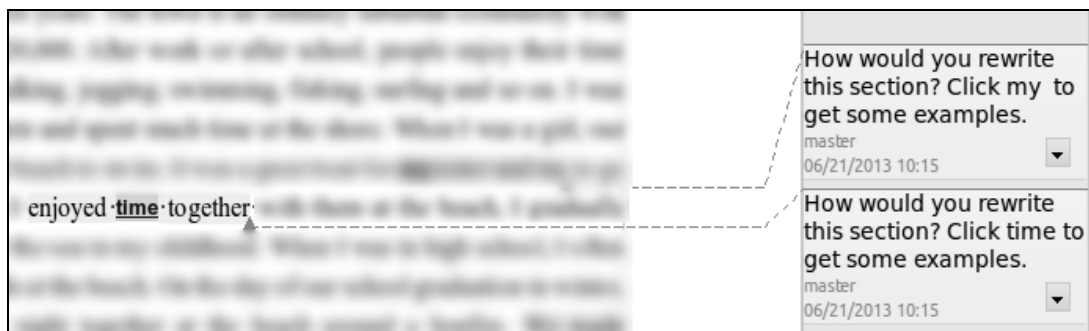


Figure 2. An example of data driven feedback.

The participant clicks the link and their browser leads them to a ready-made set of concordances, usually consisting of 5, 10 at the most 20 lines and from this they must decide where to make the correction. The participants have an option to contact the instructor to check if their revision is correct or not.

The theory behind this approach is based on the idea proposed by Gaskel and Cobb(2004) that students need vast number of examples in order to acquire the language:

> We argue that an example-based acquisition theory translates into an instructional theory only if we can do the following things: (a) vastly increase the number of examples that L2 learners are exposed to in a given unit of time, (b) organize these examples so their patterns are highlighted, (c) get learners to attend to the examples, and (d) give systematic feedback on the success of interpreting the examples. A properly configured concordance, set within a suitable learning interface, can do these things and hence can test our argument. (p.304)

While the Support Writing Course uses this approach there are some differences. Due to the time constraints (an approximately 8 week program) and the fact that the course was conducted almost entirely online rather than have regular meetings in a classroom, the instructors chose to use this approach exclusively with the learners as a mediated approach in contrast with the original study which used it to prepare them for corpus consultation. In addition at least two concordance feedback items were offered rather than five as mentioned in the Gaskell & Cobb article (ibid., p 308) in order to avoid overwhelming the participants.

**Reflections and future research**

The literature for both errors in writing instruction and for data-driven learning leave much unsaid in terms of solid data regarding effects in terms of language learning. Direct error correction might have some positive effects in terms of accuracy but the data has yet to be conclusive. Polio (2012) in her recent review of the debate on error correction, states that second language acquisition theory has not been considered when conducting research on writing feedback, and believes that by doing so the research agenda can move forward. In terms of affect, data-driven learning seems to work with learners who may enjoy the investigative aspect of the approach. Regarding its effect in terms of learning for both these approaches there needs to be more focused quantitative studies.

With this theoretical background in mind, for our writing program and its use of data-driven learning, the affective aspects of the approach should be considered. Training seems to make a difference with learners when they engage in corpora, but just showing concordances briefly is not enough. Such training must be mediated by giving learners a guide when accessing corpora. One example of such a guide is Thurston & Candlin (1998), which offers a step by step process for the learner when looking at concordances.

- LOOK at concordances for the key term and words surrounding it, thinking of meaning.
- FAMILIARIZE yourself with the patterns of language surrounding the key term by referring to the concordances as you complete the tasks.
- PRACTICE key terms without referring to the concordances.
- CREATE your own piece of writing using the terms studies to fulfill a particular function of academic writing. (p. 272)

Flowerdew (2010) commented that this guide is the only one of its kind in the literature and something similar to it has yet to be produced (p. 445).

Another aspect of data-driven learning is the question of approach as mentioned above: do we allow the participants to directly consult corpora or should the instructors mediate the output as a form of feedback? With the mediated approach students might discover the error quickly. The output is filtered by the instructor with a combination of training the students on how to look at concordances, give additional guidance given in terms of highlighting or underlining parts of the concordance to facilitate noticing the regularity or limit the output to only display one lexical or syntactic form given in the concordance.

With the direct approach learners are given free rein to use concordance software, provided there is a training period to familiarize them with reading the data. Here the correct form is to be discovered. The complexity can be minimized by consulting the teacher, or even allowing the students to work in pairs or groups when analyzing concordances.

Gilmore (2009) recommends a combined approach in order to guide his students. Since direct consultation can be complex, especially for lower level learners, allowing direct consultation also stimulates a more motivated, autonomous learning experience. His approach, like Gaskell and Cobb's is to highlight problem areas only and leave it to the learner to investigate the error. But even this can be problematic for learners for Gilmore discovered that some of his learners could not find the error (p.368). In response more direct guidance was necessary such as circling the word which was the source of error. This allowed the learner to use this word as the node word for a search and with this guidance find the correct structure in the corpus. While the learners are using corpora to find examples, the instructor can still mediate the experience through manipulating the feedback.

**Conclusion**

There has yet to be a detailed analysis of the how students respond to concordances, and what sort of training process must entail in order for it to become a resource for writing and whether it encourages learning. In terms of research opportunities in order to obtain much needed quantitative data, a replication study may be one option, but designed to take into account the logistics of our particular course.

Another avenue is more qualitative: should our program try to promote direct consultation or continue to mediate the feedback? While direct consultation seems motivating for some learners as seen above, and transfers the work of searching the

corpora to the learners, the age, tolerance of technology, and learning styles of those individuals may be factors that impact this choice.

Regardless, there are other affective issues such as introducing concordances. During one orientation session for our program, when the participants were presented with a concordance on the monitor screen, there was silence in response to one instructor's inquiry about what they can see. It should be assumed that the participants had never seen a concordance before and wouldn't know where to start. Based on that episode, the degree of mediation is vital to the presentation and use of concordances. The procedure by Thurston & Candlin (1998), mentioned above along with closer guidance might be one strategy.

Boulton (2010) reflected that the slow adoption of data-driven learning was more an issue of "implementation of DDL rather than the nature of the techniques themselves" (p 537). The issues addressed in this article are one way to meet that challenge. For our program the art to doing this may be in the metaphors we use which can facilitate understanding. One might imagine how people reacted to a microscope in the past and the resulting confusion that it ensued. The value of a microscope is because this allows the user to see details that are otherwise unseen and organize the chaos of the phenomenal world. Now it would be unthinkable to do any kind of science without such an instrument. The same can be said of a concordancer, which allows the chaos of language to be organized allowing the fact of actual language to be observed rather than rely on intuitions. Such a device is something most learners would want as a resource for writing, if only it was packaged to match their sensibilities.

REFERENCES

Abu Alsharr, A., & AbuSeileek, A. F. (2013). Using concordancing and word processing to improve EFL written English. *JALT CALL Journal*, *9*(1), 59–77.

Boulton, A. (2008). *Evaluating corpus use in language learning: State of play and future directions*. Paper presented at the Amerian Accociation of Corpus Linguisitcs, Brighham Young University.

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

Boulton, A. (2010). Data-driven learning: Taking the computer out of the equation. *Language Learning*, *60*(3), 534–572.

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo Tenorio, L. Rodríguez-Navarro, J. Santana (Eds.). *Corpora in the Foreign Language Classroom: Selected papers from the Sixth International Conference on Teaching and Language Corpora (TaLC 6).* (pp. 3-16). Kenilworth: Rodopi.

Cobb, T. (2013). Compleat lexical tutor. Retrieved July 1, 2013, from http://www.lextutor.ca/

Davies, M. (2010). The corpus of contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing*, *25*(4), 447–464.

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, *8*(1), 1–11.

Flowerdew, L. (2009). Applying corpus linguistics to pedagogy: A critical evaluation. *International Journal of Corpus Linguistics*, *14*(3), 393–417.

Flowerdew, L. (2010). Using a corpus for writing instruction. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 444–457). New York: Routledge.

Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, *32*(3), 301–319.

Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, *63*(4), 363–372.

Johns, T. (1991). Should you be persuaded: Two samples of data-driven learning materials. *English Language Research Journal*, *4*, 1–16.

McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: An advanced resource book*. New York: Routledge.

McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice.* Cambridge: Cambridge University Press.

Mull, J. (2013). Learner as Researcher: Student Concordancing and Error Correction. *Studies in Self-Access Learning Journal*, *4*(1), 43–55. Retrieved from http://sisaljournal.org/archives/mar13/mull

O'Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching.* Cambridge University Press.

Polio, C. (2012). The relevance of second language acquisition theory to the written error correction debate. *Journal of Second Language Writing*, *21*(4), 375–389.

Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, *20*(1), 83–95.

Sheppard, K. (1992). Two feedback types: Do they make a difference? *RELC Journal*, *23*(1), 103–110.

Shintani, N., & Ellis, R. (2013). The comparative effect of direct written corrective feedback and metalinguistic explanation on learners' explicit and implicit knowledge of the English indefinite article. *Journal of Second Language Writing*, *22*(3), 286–306.

Thurston, J., & Candlin, C. N. (1998). Concordancing and the teaching of the vocabulary of academic English. *English for Specific Purposes*, *17*(3), 267–280.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, *46*(2), 327–369.

Van Beuningen, C. G., De Jong, N. H., & Kuiken, F. (2012). Evidence on the effectiveness of comprehensive error correction in second language writing: Effectiveness of comprehensive CF. *Language Learning*, *62*(1), 1–41.

Widdowson, H. G. (1991). The description and prescription of language. *Georgetown University Round Table on Languages and Linguistics*, *1991*, 11-24.

Yoon, H., & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of Second Language Writing*, *13*(4), 257–283.