

Towards the pedagogical application of corpora:
A case study on data-driven learning

Eri Hirata

Fukuoka Jo Gakuin University Bulletin.
Faculty of International Career Development

福岡女学院大学紀要

国際キャリア学部編抜刷 Vol. 10, 2024

Towards the pedagogical application of corpora: A case study on data-driven learning

Eri Hirata

Introduction

Corpus linguistics has influenced many areas of linguistics and applied linguistics (Hunston, 2022). Among them, it has made a great impact on the creation of pedagogical materials used in English language teaching (ELT) and learning. The most significant example of such impact can be found in lexicography, the creation of corpus-based language dictionaries (e.g., *COBUILD - the Collins Birmingham University International Language Database*) and other reference materials, such as grammar reference books (e.g., *Longman Grammar of Spoken and Written English*, Biber, et al, 1999). Grammar teaching materials (e.g., *Real Grammar* by Conrad and Biber, 2009), and the ELT materials are also created by exploiting corpora. For instance, *Touchstone* (McCarthy, McCarten and Standiford, 2005; 2024) is a ‘corpus-informed’ course which includes dialogues derived from corpora. Moreover, investigating learner corpora, which consist of language produced by learners, can inform “the understanding of acquisition, and the improvement of pedagogic materials” (Hunston, 2022: 143).

There is also a direct application of corpora in the classroom as a language pedagogy, termed “Data-driven learning (DDL)” (Johns, 1991), which involves:

the use in the classroom of computer-generated concordances to get students to explore regularities of patterning in the target language and the development of activities and exercises based on concordance output (Johns and King 1991: iii).

There have been a number of research projects done by applied corpus linguists in this area and a wide range of corpus-based DDL activities have been proposed. However, the challenges remain and it cannot be said that DDL is a widely used form of pedagogy in ELT today. Despite the contribution of corpus linguistics to ELT and the improvement of pedagogical materials, learners and sometimes even practitioners are less

informed of what the information on usages or examples in the materials is based on. Indeed, if practitioners themselves do not know about what corpora can offer, it would be impossible to incorporate any corpus-based exercises in the classroom.

As Römer (2009) notes, for the successful integration of corpus-based exercises and DDL in language learning contexts, it is necessary to train teachers in working with corpora. Considering the busy schedule of practitioners in general, Römer (2009: 92) suggests that such 'corpus mission' can start early by incorporating the elements of corpus linguistics within the training when trainees are at universities. Nevertheless, little is known in terms of how such an incorporation of corpus-based activities is perceived by learners who are new to corpora at the undergraduate level.

This study attempts to explore students' attitudes towards the incorporation of corpus-based activities within a content course on ELT and material designs offered for undergraduate students majoring in English language at a university in western Japan. The course reported in this study included some instructions on corpus linguistics and hands-on sessions on using corpora. After five weeks of dealing with corpus-related topics and DDL activities, a survey was conducted with the participants.

The following section first reviews the advantages and challenges of the direct application of corpora in ELT. This is followed by an explanation of the study procedures and DDL activities that the participants experienced. Survey findings will be reported with a discussion of the benefits and challenges of incorporating DDL in such a course, and further studies will also be discussed.

Direct application of Corpora in ELT: Data-driven Learning (DDL)

As mentioned above, in data-driven learning (DDL), learners are encouraged to consult a corpus for language usage by themselves. The 'father' of DDL, Tim Johns (1997: 101) explains the DDL approach using some metaphors:

The central metaphors embodying the approach are those of the learner as a 'linguistic researcher' testing and revising hypotheses, or as a 'language detective' learning to recognise and interpret clues from context ('Every student a Sherlock Homes').

In other words, DDL encourages learners to learn in an inductive manner,

through the observation of concordance outputs. As Hunston (2022: 1) notes, “[w]hat distinguishes a corpus from a collection of digitised texts is that it is formatted such that the application of the software enables patterning to be observed that would be missed by conventional forms of reading”. This experience of observing concordances often encourages ‘noticing’ (Schmidt, 1990) in terms of language use (e.g., patterns, collocations), and such a process is likely to contribute to increasing learners’ lexical and grammatical knowledge and promote their language awareness.

Through consulting a corpus directly, learners are encouraged to explore the use of certain items autonomously and to learn by discovering (Bernardini, 2004). Therefore, by participating in DDL, learners not only benefit in terms of expanding their knowledge of the language but also become autonomous through consulting the corpora by themselves (Godwin-Jones, 2017).

Moreover, learners can be exposed to the authentic use of language through the analysis of corpora in DDL. General corpora usually consist of language output, both spoken and written, produced by native speakers. In a sense, learners can access authentic examples of how native speakers use the language both in spoken and written English through corpora. Barnbrook (1996: 140) described the corpus as “a tireless native-speaker informant, with rather greater potential knowledge of the language than the average native speaker”. One might argue that having search engines on the Internet might do the same work these days, but one of the big differences is that the general corpora are designed to represent a certain language, considering what kind of texts to be included in the corpus in terms of their balance, size, representativeness (Biber, 1993) in the process of their compilation. Although there remain some ongoing discussions on the issue of representativeness (e.g., Egbert, Biber & Gray, 2022), the corpora are certainly more reliable sources than the Internet when checking for language usages.

Despite the advantages of the direct use of corpora in pedagogy, there are several challenges also identified in the literature. One of the obvious challenges is related to the format of the language outputs in Key-Word-In-Context (KWIC) or concordances. The corpus software usually presents language different from what learners are used to. This is one of the benefits of corpora, which allows users to notice patterns, as mentioned above. However, some novice users may find it challenging to use the

corpora at first until they get used to the concordance outputs. As Hunston (2022) notes, “[o]bserving pattern is ... easiest when concordance lines are presented in KWIC format, but this format can be difficult for learners to process” (Hunston, 2022: 152). Boulton and Cobb (2017: 351) also point out:

Chopped-off concordance lines may help expose patterns yet be off-putting to some and are not designed for gaining meaning as traditionally conceived via linear reading; most corpora are composed of authentic native language well beyond the comfort level of many learners; and DDL work requires substantial training, and the processes are time consuming when learners could simply be told or use pedagogically derived resources such as dictionaries.

Moreover, in addition to the presentation of the data in concordance format, the amount of the language data presented can be another challenge for learners. Osborne (2004: 252) also mentions, “unless corpus examples are filtered in some way ... many of the contexts are likely to be linguistically and culturally bewildering for the language learners”.

Nevertheless, recent learners who can be considered as ‘digital natives’ who have grown up surrounded by digital materials, may respond to corpora differently. As Boulton (2012: 25) also argues:

It seems likely that many learners around the world are already Googling the Internet in ways not entirely dissimilar to DDL, a practice which may be actively encouraged by their teachers while remaining invisible in the DDL research literature.

Indeed, some learners might find it relatively easy to work on DDL as the experience itself is similar to searching on the Internet, but there may be some who might still find it challenging because of their unique form of presentation (i.e., KWIC) and amount of data, as pointed out earlier. Nevertheless, it can be argued that it is worth exploring the possibilities of the application of corpus-based activities and DDL, as the advantages are likely to outweigh the challenges mentioned above.

For successful corpus-based learning, Mukherjee (2006: 14) suggests that learners should have some ‘corpus literacy’, which includes the following four points:

- 1) basic understanding of what a corpus is;
- 2) knowing what can and cannot be done with a corpus;

- 3) being able to analyse concordance lines; and
- 4) understanding how to draw conclusions based on corpus data about language trends and use.

The benefits of having some ‘corpus literacy’ have been acknowledged and incorporating corpus-related activities in teacher training courses has also been suggested in the literature (e.g., Farr, 2008; Heather and Helt, 2012, Mukherjee, 2004, Römer 2009). If learners have the opportunity to be introduced to data-driven learning (DDL) at university, it might be possible to acquire some ‘corpus literacy’ through participating in the activities. With this in mind, the course reported in this study included some instructions on corpus linguistics and hands-on sessions on using corpora. In the following section, details of the procedure and methods of the case study will be presented.

The Case Study: Procedure and Methods

In this study, hands-on DDL tasks were introduced within a content course on ELT and material design, offered for undergraduate students majoring in English Language, during the second semester in the academic year 2023. Five weeks (90-minute class per week) were spent on both in-class and out-of-class DDL activities and follow-up tasks were given as homework each week. Two types of DDL have been tried in this study: 1) a “deductive DDL” which is a teacher-mediated, and considered a “soft” approach, as opposed to 2) an “inductive DDL” which is a “hard” type which requires learners to be more autonomous and work on their own (Cresswell, 2007: 279). The DDL approach incorporated in this study have mainly three stages:

- Stage 1: Introduction to corpus linguistics & Deductive DDL tasks
- Stage 2: Inductive DDL tasks
- Stage 3: Post-DDL tasks: Presentation & Questionnaire feedback

In Stage 1, as an introductory session, basic background information of corpus linguistics was introduced with some hands-on activities using different tools and corpora in order to familiarise participants with some key issues such as frequency, and identifying patterns in KWIC (Key-Word-In-Context) format. Computer software, *AntConc 4.2.4* (2021) developed by Laurence Anthony and *Wmatrix 5* (2023) by Paul Rayson

were used to show how self-compiled corpora or electronic texts can be analysed using concordance software tools.

Two types of corpora, Corpus of Contemporary American English (COCA) and British National Corpus (BNC) in English-Corpora.org, were mainly used for DDL tasks. After each student signed up for using English-Corpora in the first week, students participated in teacher-led 'deductive DDL' tasks in a computer lab together with the instructor in the second and third weeks in class. In these sessions, the instructor mainly introduced different query syntax and wildcards (*word, anyword, POS searches, synonyms, collocations* etc.), to make students familiarise themselves with the different ways of analysing corpora. These teacher-led deductive DDL tasks were mainly used as opportunities for the participants to get used to using the tools and discuss the findings with their peers in class. Throughout the in-class sessions, support was provided for participants on the hands-on issues, when necessary.

Based on the training received in Stage 1, in Stage 2, participants worked on an "inductive DDL" task in which they were asked to decide their own query syntax and keep the record on a worksheet of their queries, along with the explanation of the initial reasons for their queries and findings. The first part of the inductive DDL was done in class so that participants could ask questions, if necessary. After they worked on some inductive DDL activities in class, the remaining tasks were given as homework to be completed on their own.

In Stage 3, as post-DDL tasks, the participants were asked to give a presentation on their findings of the inductive DDL in front of their peers and give feedback to each other. After that, they were asked to complete the questionnaire survey reflecting on their experience with the corpora during the course. In the next section, the result of the survey will be reported and discussed.

Survey Results & Discussion

As a post-DDL task, the questionnaire survey was conducted and feedback was collected from the participants in order to understand their perceptions on the DDL tasks introduced to them during the course. For the analysis, the responses that had participants' permission to use them for the purpose of this research were included, which yielded an overall response rate of 95.8 per cent (n=23). The participants were all non-native

speakers of English (Japanese as their L1), majoring in English Language at a university in western Japan, and enrolled in a content course related to ELT and material design. They had averaged 9.8 years of learning English. The questionnaire covered topics mainly related to the reflections on DDL tasks and their opinions about them.

Regarding the participants' prior experience of using corpora (Item 3.1), it was found that none of them had any experience in using corpora. In response to Item 3.2, which asked whether the participants knew about corpora before participating in this course, 17.9 per cent of the participants answered they had heard of corpora before, but most of them (i.e., 82.6 per cent) reported that they did not even know about corpora. It can be summarized that all the participants were new to corpora, especially with no prior hands-on experience of them.

For the participants' experience of DDL and the use of corpora, the survey included statements to which participants were asked to agree or disagree by choosing one of the options from "I fully agree", "I partly agree", "I partly disagree", and "I completely disagree". The result of the responses is summarized in Table 1. As can be seen, while 65.2 per cent of the participants reported that it was not easy to use the corpora, 34.7 per cent of them found it easy to use the corpora (Item 4.1, $M=2.17$, $SD=.70$).

The survey also asked about the participants' ICT ability, their general experience of using PC and their attitudes towards using any new tools or software. In order to see if there are any relationships between the participants' ICT ability and their experience and attitudes towards corpora, Pearson conduct-moment correlation coefficients were calculated. There was a positive correlation between Item 2.1 and Item 4.1 ($n=23$, $r=.46$, $p<.05$), which indicates that the participants who are generally confident in using computers were likely to have found the use of corpora easy. In Hirata (2020), it was found that there was a tendency for participants who reported lacking confidence in their ICT skills, to disagree with the usefulness of general corpora for their study or future use (ibid: 99). In this study, however, there were no notable relationships observed with the participants. It can be assumed that participants in this study had experienced online classes during the years affected by the Covid-19 pandemic, and may have got used to using digital materials (cf. Table 1: 65.2 per cent of the participants either fully or partly agreed to Items 2.1; $M=2.70$, $SD=.80$ and Item 2.2, $M=2.61$, $SD=.71$). Moreover, in terms of the usefulness of the corpora (Item 4.2; $M= 3.22$ $SD=.59$), the

Table 1: Descriptive statistics for the items focusing on participants' experience of ICT & DDL

No.	Item	M	SD	Response (%)			
				1. completely disagree.	2. partly disagree.	3. partly agree.	4. fully agree.
2.1	I regularly use PC and I am confident of its use.	2.70	.80	8.70	26.90	52.17	13.04
2.2	I usually feel comfortable using new tools, software and apps.	2.61	.71	8.70	20.09	60.87	4.35
4.1	Using the corpus was easy.	2.17	.70	17.39	47.83	34.78	0.00
4.2	It was useful to check the usages through the corpus.	3.22	.59	0.00	8.70	60.87	30.43
4.3	I think using the corpus will help improve my English.	2.83	.70	0.00	34.78	46.83	17.39
4.4	The corpus can be used (applied) in ELT.	2.74	.53	0.00	30.43	65.22	4.35
4.5	I would like to use the corpus for checking usages for my own study in future.	2.96	.62	0.00	21.74	60.87	17.39

n=23

majority of the participants (i.e., 91.3 percent) either fully or partly agreed to the statement. It shows that the participants recognized the usefulness of the corpus for checking usages, regardless of how they felt towards the tasks.

As noted earlier, the participants experienced mainly two types of DDL tasks in this study: 1) teacher-led “deductive DDL” tasks in which the instructor showed how/what they could search using corpora by sharing the instructors’ screen on the monitors in the classroom; and 2) “inductive DDL” in which they were asked to decide their query syntax and record the findings of the task, along with the explanations of initial reasons of their queries. The survey also included a question to ask about the participants’ preference over deductive or inductive DDL tasks, which was followed by a comment section which enabled the participants to provide their reasons for their choice (Item 4.7).

56.5 per cent of the participants answered that they preferred the teacher-led, deductive DDL tasks, and for the reasons, some left comments

such as below:

- *“Because it is easier to search if the search terms are specified by the teacher.” (No.1).*
- *“I found it hard to think of search terms on my own.” (No.2)*
- *“Because you do not have to look up words that don’t come up in search results by mistake. (No.7)*
- *“When I did it on my own, there were many search results and I felt I wasn’t making good use of them.” (No.22)*

As can be seen, comments were observed reporting that the participants found it challenging to think of the query syntax by themselves or work on the task inductively. These results align with the findings in the existing research (e.g., Boulton, 2009; Crosthwaite, 2017) in the sense that it usually takes some time until learners get used to consulting the corpora in DDL. The inductive nature of the task might have raised concerns about whether they were doing it right. This might be related to certain learning styles each participant is comfortable with. The comments above seem to suggest that for those students, there may be a need for some more extra hands-on instructions before moving on to inductive DDL tasks, or having some query options or suggestions might have helped them to start with.

Still, it is important to note that not everyone preferred the deductive DDL tasks. In fact, 43.5 per cent of the participants showed their preference for the inductive DDL. As for the reasons for preferring inductive DDL, for instance, some reported:

- *“I learned a lot by being able to search for the words and phrases that I was interested in and learning about them in detail from the examples.” (No.11).*
- *“If you decide your query, you can also compare the results and find out what you want to know through various examples of usages.” (No.12)*
- *“I was able to look up words that I was interested in or wanted to look up so it was better for me to think of queries freely.” (No.13).*
- *“There were times when I looked up a certain word’s usage and I thought, ‘Ah, I see!’ because I was interested in the usages of the particular item. I thought it is likely to contribute to long-term memory.” (No.16)*
- *“I was able to learn a lot by researching English words that interests me.” (No.17).*

For many of the participants who preferred the inductive DDL, the most frequently reported reason was “being able to decide what words/phrases to search based on their interest”. Such attitudes show a sign of having autonomy, which is one of the key factors for benefiting from DDL, because of its inductive nature (Chambers, 2005).

In order to see if there were any relationships between items related to the participants’ attitudes towards the use of corpora, and the participants’ preference for the types of DDL (i.e., either deductive DDL or inductive DDL), the differences in the mean scores and their p-values were calculated for each item. As can be seen in Table 2, the data suggest that participants who preferred inductive DDL tended to answer positively for Item 4.3 ($M= 3.20$ $S.D.=.40$), which is related to the use of corpora for their study. Moreover, it was found that the participants who preferred inductive DDL were likely to answer positively about the possible application in ELT (i.e., Item 4.4; $M= 3.00$ $S.D=.45$), compared to those who preferred the teacher-led deductive DDL.

Table 2: Relationship between participants’ attitudes and the types of DDL

No.	Item	Item 4.7				P
		1. Deductive DDL		2. Inductive DDL		
		M	SD	M	SD	
4.1	Using the corpus was easy.	2.00	.68	2.40	.66	.19
4.2	It was useful to check the usages through the corpus.	3.15	.66	3.30	.46	.56
4.3	I think using the corpus will help improve my English.	2.54	.75	3.20	.40	.02*
4.4	The corpus can be used (applied) in ELT.	2.54	.50	3.00	.45	.04*
4.5	I would like to use the corpus for checking usages for my own study in future.	2.77	.58	3.20	.60	.11

$n=23$

* $p<.05$

The survey also included open-ended questions which asked participants to write about their honest opinions about the incorporation of DDL tasks in the course. There are some comments which indicate that the benefits of corpora had been recognised by the participants:

Towards the pedagogical application of corpora:
A case study on data-driven learning

- *“I had never heard of corpora before and I used it for the first time. I am glad to know that I can learn about English through corpora. I would like to continue using it whenever I have the opportunity.” (No.11)*
- *“I was able to find out if I was using certain words or phrases correctly by accessing corpora. Also, I was able to learn the new expressions that I didn’t know before.” (No.9).*
- *“We can expand your knowledge of English through the huge number of examples, rather than just being limited to your own knowledge.” (No.16).*
- *“I think it will help increase my vocabulary knowledge because we can check the relationship between words and whether they can be used together.” (No.20).*
- *“When you feel like you are using the same expressions, you can look for synonyms or similar expressions for alternatives.” (No.22).*

Moreover, as can be seen in the following, some positive comments were found within the participants’ responses, even when they start with their comments mentioning the challenges involved in using the corpora. These seem to indicate that there is still room for hope that the participants might make use of the corpora under the suitable circumstances in the future:

- *“It was difficult to use the corpus as I am not used to it. But I think it’s especially useful when looking up words or expressions that can be used in writing.” (No.3).*
- *“It was hard to remember how to use wildcards and I struggled with the tasks. I think the corpus is useful in checking how to use newly learned words or phrases, and to improve my vocabulary knowledge.” (No.10).*
- *“Although there were times, I was not sure how to search and it was not easy to use, but I think it would be a very useful teaching material once you master how to use it.” (No.12)*
- *“At first, I was confused because some errors occurred when working with the corpus. But I gradually found it useful to discover the various examples of word usages.” (No.15).*
- *“I am not good at using computers, so it was a little difficult for me to use it. However, I was able to learn how to use certain words in detail and I thought it would be useful for my learning.” (No.16).*

In terms of future use, a somewhat unenthusiastic comment was also observed:

- *“To be honest, I don’t think I would use a corpus because there are so many different ways to study English. And it was a bit difficult for me. However, I was able to access a lot of example sentences and their frequency information, which led to new discoveries for me. I learned a lot thanks to the corpus tasks.” (No.20).*

The above was not entirely surprising, and many studies on DDL report similar findings which showed resistance towards DDL because of the amount of data and difficulties associated with arriving at any conclusions (e.g., Boulton and Cobb, 2017; Osborne 2004). However, it is also important to remember what Boulton (2009: 10) notes:

Learners, like teachers, might find the messy nature of real language in use to be destabilising at first, preferring the teacher to have all the answers. But it would seem disingenuous to coddle learners with simplified language, disempowering them and leaving them unprepared for the realities of the authentic language we are presumably preparing them for.

Indeed, the fact that none of the participants had the experience of using corpora even after an averaged 9.8 years of learning English, shows that the corpora have not yet been widely implemented in the Japanese teaching context, as the following comment illustrates:

- *“I think there are many English learners, like myself, who don’t know that such useful tools exist. So, I am very glad that corpus activities were incorporated into this course. (No.11).*

As mentioned by several participants, incorporating the DDL tasks in the current course provided the first opportunity for them to be acquainted with corpora and be introduced to the new way of learning English through DDL. In particular, consulting corpora, as the participants experienced during the inductive DDL tasks, can be added as a way of their autonomous language learning, or even as one of the “discovery strategies” (Schmitt, 1997) that they could employ when they come across any new words or phrases in their learning.

Conclusion and Further Study

This study explored students' perspectives on the use of corpora within a content course on ELT at a Japanese university. The survey revealed that participants needed training to get used to using the corpora, but most of them were able to make use of them for their enquiries in the DDL tasks after receiving some instructions. As reported, just over half of the participants answered that they preferred teacher-led, deductive DDL activities over the inductive type. Since some comments who preferred deductive DDL showed their uncertainty towards deciding their own searches, it was suggested that some additional instructions (e.g., presenting some options for them to choose from) might have helped those who found it challenging to handle the corpora, before moving on to the inductive DDL tasks. This may help learners to come up with their own searches with more variations.

At the same time, it was significant that most of the participants, including those who found using the corpora challenging, recognized the usefulness of corpora (cf. Item 4.2; $M=3.22$ $SD=.59$). Moreover, one of the encouraging results was that nearly half of the participants preferred inductive DDL, and to explore the corpora on their own by deciding their own query syntax, after receiving in-class instructions (cf. Item 4.7). Such attitudes indicate that learners can be more autonomous users of corpora, which is likely to assist in promoting their language awareness through corpus analysis.

While the results do not seem to show any definitive conclusions, they indicate that incorporating the hands-on instructions of corpora use and DDL tasks in a content course was meaningful in terms of introducing the corpora to the participants for the first time and convincing them about their usefulness. Though the challenge remains with the difficulties involved with using corpora, several participants noted that they feel they can make use of a corpus, once they are familiarised with its use.

Nonetheless, the results presented in this study should be seen in light of some limitations. One of the obvious limitations is the relatively small number of participants, and the result of this study may not be easily generalised. Moreover, due to time constraints, it was not feasible to have interviews with individual participants. Though the survey included open-ended questions to which participants provided their honest opinions and interesting insights, having follow-up interview sessions with the

participants might have offered further information about their perceptions of the DDL tasks incorporated in the course. This could be addressed in a future study. Indeed, more studies are needed in order to fully understand the learners' perspectives toward DDL, and further research on the efficacy of DDL may also be necessary. Though the number of participants may be small at each time, doing some case studies like the present one at the tertiary level could be one of the ways to invite learners to experience what corpora can offer. It is hoped that participants who were convinced of the usefulness of having access to corpora will continue to make use of them for their studies and try to incorporate corpus-based activities when/if they go into teaching in the future.

References

- Anthony, L. (2021). *AntConc*. 4.2.4 [Computer Software] Tokyo: Japan: Waseda University, Available from: <https://www.laurenceanthony.net/software.html>
- Barnbrook, G. (1996). *Language and Computers. A practical introduction to the computer analysis of language*. Edinburgh: Edinburgh University Press.
- Bernardini, S. (2004). Corpora in the classroom: an overview and some reflections on future developments. In J. H. Sinclair (Ed.), *How to Use Corpora in Language Teaching*. (pp.15-36). Amsterdam: John Benjamins.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing* 8(4), 243-257.
- Biber, D., S. Johansson, G. Leech, S. Conrad & E. Finegan (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Boulton, A. (2009). Data-driven learning: Reasonable fears and rational reassurance. *Indian Journal of Applied Linguistics*, 35(1), 81-106.
- Boulton, A. (2012). What data for data-driven Learning? *EuroCALL Review*, 23-27.
- Boulton, A. and Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393.
- Chambers, A. (2005). Integrating corpus consultation in language studies. *Language Learning & Technology* 9(2), 111-125.
- Conrad, S. and Biber, D. (2009). *Real Grammar: Corpus-based approach to English*. Pearson Longman.
- Cresswell, A. (2007). Getting to 'know' connectors? Evaluating data-driven learning in a writing skills course. In E. Hidalgo, L. Querada & J. Santana(eds.), *Corpora in the Foreign Language Classroom* (pp. 267-287). Amsterdam: Rodopi.
- Crosthwaite, P. (2017) Retesting the limits of data-driven learning: feedback and error correction, *Computer Assisted Language Learning*, 30(6), 447-473.
- Davies, M. (2008-) *The Corpus of Contemporary American English (COCA)*. Available from <https://www.english-corpora.org/coca/>.
- Davies, M. (2004) *British National Corpus (from Oxford University Press)*. Available from <https://www.english-corpora.org/bnc/>.

Towards the pedagogical application of corpora: A case study on data-driven learning

- Egbert, J. D. Biber & B. Gray. (2022). *Designing and evaluating language corpora*. Cambridge: Cambridge University Press.
- Farr, F. (2008). Evaluating the use of corpus-based instruction in a language teacher education context: Perspectives from the users. *Language Awareness* 17(1), 25-43.
- Gabrielatos, C. (2005). Corpora and language teaching: Just a fling or wedding bells? *Teaching English as a Second Language. Electronic Journal*, 8/4, 1-35. Available online: <http://tesl-ej.org/ej32/a1.html>. [2023/12/22].
- Godwin-Jones, R. (2017). Data-informed language learning. *Language Learning & Technology* 21, 9-27.
- Heather, J. and Helt, M. (2012) 'Evaluating corpus literacy training for pre-service language teachers: Six case studies.' *The Journal of Technology and Teacher Education* 20(4), 415-440.
- Hirata, E. (2020). The development of a multi-modal corpus tool for young EFL learners: a case study on the integration of DDL in teacher education. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for Pre-tertiary Learners*. (pp. 88-105). Oxon: Routledge.
- Hunston, S. (2022). *Corpora in Applied Linguistics*. 2nd edn. Cambridge: Cambridge University Press.
- Johns, T. F. (1991). Should you be persuaded - Two samples of data-driven learning materials. In T.F. Johns & P. King (eds.), *Concordancing English Language Research Journal* 4, 1-16.
- Johns, T. F. (1997). 'Contexts: the background, development and trialling of a concordance-based CALL program.' In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and Language Corpora*. (pp. 100-115). Harlow: Addison Wesley Longman.
- Johns, T. F. and King, P. (1991). From printout to handout: Grammar and vocabulary teaching in the context of data-driven learning. *Classroom Concordancing English Language Research Journal* 4, 27-45.
- McCarthy, M., McCarten, J. and Sandiford, H. (2005). *Touchstone*. Cambridge: Cambridge University Press.
- McCarthy, M., McCarten, J. and Sandiford, H. (2014). *Touchstone*. 2nd edn. Cambridge: Cambridge University press.
- Mukherjee, M. (2004). Bringing the gap between applied corpus linguistics and the reality of English language teaching in Germany. In U. Connor & T. A. Upton (Eds.), *Applied Corpus Linguistics: A multi-dimensional perspective*. (pp.239-250). Amsterdam: Rodopi.
- Mukherjee, J. (2006). 'Corpus linguistics and language pedagogy: the state of the art - and beyond' In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. (pp. 5-24). Frankfurt am Main: Peter Lang.
- Rayson, P. (2023). *Wmatrix 5*. [Computer Software] Available from <https://ucrell.lancs.ac.uk/wmatrix/>
- Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? In K. Aijmer (Ed.), *Corpora and language teaching*. (pp.81-98). Amsterdam: John Benjamins.

Eri Hirata

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, and pedagogy* (pp. 199-227). Cambridge: Cambridge University Press.