

Norwegian in-service teachers' perspectives on language corpora in teaching English

Barry Kavanagh

Inland Norway University of Applied Sciences

Abstract

This study aims to explore potential reasons why the use of the tools and methods of corpus linguistics are not prevalent in English teaching in Norway, using the research question What do in-service English teachers in Norway find useful about corpora and what do they find challenging? The study provides interview data from in-service teachers, contributing to our understanding of the in-service perspective on corpora. The research design consists of teaching corpus use in seminars for in-service English teachers (featuring LancsLex, the concordancer AntConc and the OANC), integrated into a language course that is part of a further education programme, and semi-structured interviews with four of the students who took the course, during which they also interacted with Netspeak, SKELL and COCA. As with previous research, the in-service teachers found corpora particularly useful for teaching and learning vocabulary, and found challenges to use which are categorized here as usability (criticism of AntConc), IT challenges (a lack of IT skills among teachers), learner-corpus interaction challenges (the complexity of software and concordance lines for pupils; pupil uninterest in language), and lack of teacher need (mistakes being “obvious” to teachers in the lower years). The article discusses some implications of these findings.

Keywords: *English language teaching, pedagogical corpus application, corpora*

Introduction

Although research has shown that corpora are useful for English language learning (Boulton & Cobb, 2017), and there is potential for using corpora in schools (Braun, 2007; Crosthwaite, 2020), it seems that the tools and methods of corpus linguistics are not directly pedagogically applied¹ very often by English teachers in Norwegian schools (Kavanagh, 2021). It has been recognized that an understanding of the teacher perspective is essential for corpus linguists to understand pedagogical needs (Braun, 2007, p. 326; Römer, 2009, p. 83). This interview-based study focuses on the views of in-service² English teachers in Norway on corpora. This adds to our scant knowledge of the perspective on corpora of in-service teachers. The research question is *What do in-service English teachers in Norway find useful about corpora and what do they find challenging?* This question focuses on the “useful” and the “challenging” as a manner of exploring potential reasons why corpus use is not more prevalent. What in-service teachers consider useful about corpora might reveal the extent of the relevance of corpora to them, and what they consider challenging may indicate why

¹ The distinction is made between “direct applications” of corpora (the use by teachers and/or pupils) and “indirect applications” (affecting what goes into reference books, textbooks, and syllabi), following Römer (2011, p. 207).

² In-service: “a teacher that has certification or is already teaching in a classroom, in contrast to a preservice teacher, who is in the process of preparing to become a teacher” (Koellner & Greenblatt, 2018).

corpora are not widely used. All of this contributes to an understanding of the teacher perspective.

Previous research is discussed in the next section, which is followed by a section on methods and a section on results. The final section is a discussion with some concluding remarks.

Previous research

Research has elucidated perspectives of language teacher educators (Breyer, 2011, pp. 117-154), and higher education teachers (Lin & Lee, 2015) on corpora, but most relevant studies have small numbers of pre-service teachers as informants, either exploring informants' perspectives on teaching with corpora (Breyer, 2011; Leńko-Szymańska, 2014; Zareva, 2017), or assessing (Leńko-Szymańska, 2017) or getting informants to evaluate (Ebrahimi & Faghih, 2016) the usefulness of corpus instruction. This study complements the above by providing interview data from in-service teachers. The in-service teacher perspective is not entirely unexplored: in some cases, there are in-service teacher informants mixed in with pre-service teachers (Ebrahimi & Faghih, 2016; Zareva, 2017), and perspectives of specifically in-service teachers are occasionally the focus (Karlsen & Monsen 2020; Kavanagh, 2021). The latter studies included teachers who were previously educated in corpus use (Karlsen & Monsen, 2020), or who directly apply corpora in their teaching (Kavanagh, 2021). The present study's informants differ, in that they were not already using corpora at the onset of the study.

Comparison between this study's data and previous studies is assisted by creating broad categories for the perceived usefulness of corpora and the perceived challenges in the use of corpora. The categories are not intended to be exhaustive or mutually exclusive. The categories are a way of sorting common positive and negative statements about direct pedagogical corpus applications found in the previous research.

Four categories of perceived usefulness are used here, from the perceptions of informants when making positive statements about corpora, for example when asked of the advantages of having corpus literacy skills (Zareva, 2017, p. 75).

Teachers' language awareness: teacher educator and pre-service teacher informants perceived teacher-corpus interaction as increasing pre-service teachers' language awareness (Breyer, 2011, pp. 149, 206; Zareva, 2017, p. 75), by leading them to "reflect on language use, their own knowledge of a specific linguistic item, textbook versus authentic language use..." and so on (Breyer, 2011, p. 206).

Vocabulary. Corpora were perceived as useful in the teaching and learning of vocabulary, by pre-service teachers (Ebrahimi & Faghih, 2016, p. 128; Leńko-Szymańska, 2014, p. 271; 2017, p. 233) and in-service teachers (Kavanagh, 2021). Informants have not always specified how. Pre-service teachers have retrieved new vocabulary from corpora for pupils to practice (Leńko-Szymańska, 2017, p. 233), and in-service teachers have detailed how they use collocation, idiom and frequency in teacher-corpus interaction and how they use collocation, idiom and checking acceptability in pupil-corpus interaction (Kavanagh, 2021, pp. 13-16).

Authenticity. Corpora were perceived as a source of examples of authentic language by pre-service teachers (Breyer, 2011, pp. 205-206; Ebrahimi & Faghih, 2016, p. 128; Leńko-Szymańska, 2017, p. 233; Zareva, 2017, p. 75). An authentic text can be defined as "a stretch of real language, produced by a real speaker or writer for a real audience and designed to convey a real message of some sort" (Morrow, 1977, p. 13, cited in Breyer, 2011, pp. 60-61).

The use of authentic language is encouraged in language teaching (e.g. Mishan, 2005, pp. 21-43), and there is a requirement to include it in the English subject in Norway (Utdanningsdirektoratet, 2019, pp. 2-7). It is usual for language in a corpus to be described as authentic (Boulton & Cobb, 2017, pp. 349, 350; Römer 2009, p. 90; Römer 2011, pp. 209, 210), although not uncontroversially (Breyer 2011, pp. 89-90; Widdowson, 2000).

Benefits of learner-corpus interaction. Learners in control of their learning was perceived as an advantage of corpora by pre-service teachers (Ebrahimi & Faghieh, 2016, p. 128). Higher education teachers perceived the transformation of students into active learners through corpus work in that learners “started to think about what they were learning” instead of just memorizing rules (Lin & Lee, 2015, pp. 269-270).

Five categories of perceived challenges are used here, from the perceptions of informants when making statements about obstacles to the use of corpora, for example when asked to express the “cons of applying CL [corpus linguistics] in ELT [English language teaching]” (Ebrahimi & Faghieh, 2016, p. 128).³

Usability. Pre-service teachers perceived corpus software as difficult to use (Ebrahimi & Faghieh, 2016, p. 128; Zareva, 2017, p. 75), or hard to remember how to use (Leńko-Szymańska, 2014, p. 269), or they lacked confidence with it (Leńko-Szymańska, 2014, p. 271). In-service teachers thought it too difficult for pupils to use (Karlsen & Monsen, 2020, p. 131; Kavanagh, 2021, pp. 17-18).

Computer and IT challenges. Teacher educators (Breyer, 2011, p. 150) and some pre-service teachers (Ebrahimi & Faghieh, 2016, p. 128) perceived a lack of computer skills among teachers, and inadequate IT infrastructure in schools.

Learner-corpus interaction challenges. Perceived obstacles to learner-corpus interaction include: classes too short in duration for work with corpora, according to both teacher educators (Breyer, 2011, p. 150) and higher education teachers (Lin & Lee, 2015, p. 271); pre-service teachers’ reluctance to use data-driven learning (DDL) with pupils, possibly because the teachers “had not mastered the expertise in designing hands-on DDL activities” (Leńko-Szymańska, 2017, p. 233); and the perception of some pre-service teachers that corpus work requires inductive learning, which lacks appeal for some pupils (Ebrahimi & Faghieh, 2016, p. 128).⁴

Lack of teacher need. Corpus-using in-service teachers said they did not need to use corpora in contexts where they were already familiar with common pupil mistakes (Kavanagh, 2021, p. 18). Also, the higher years of the English subject are more topic-focused than language-focused, according to in-service teachers (Kavanagh, 2021, p. 17), which may reduce the need for linguistic data.

Workload. Difficulty of adding to teachers’ workloads, in terms of the time available for activities in class, was expressed by teacher educators (Breyer, 2011, p. 150) and higher education teachers (Lin & Lee, 2015, p. 271). The latter teachers also saw preparing teaching materials from corpus data as a workload problem (Lin & Lee, 2015, p. 270-271).

³ A challenge to corpus use that is not categorized here, but of which some in-service teachers are aware, is the financial cost of accessing corpora (Karlsen & Monsen, 2020, p. 134). It is not categorized because the informants of the present study had no problems accessing the corpora they worked with, thus the topic would not be covered in the interviews.

⁴ On deductive corpus-based learning, see for example Liu & Lei, 2017, pp. 31-34.

Methods

As the in-service teachers were not corpus users, it was decided to provide them with introductory corpus seminars, and afterwards conduct interviews. The seminars are thus the contextual backdrop of the interviews. This study's interview data add to our knowledge of the perspective on corpora of in-service teachers. Also, the interview data contribute to an area of research in which there are few face-to-face interviews on the topic of corpora with any kind of teacher (Breyer, 2011, pp. 117-154; Karlsen & Monsen, 2020; Kavanagh, 2021). The purpose of using semi-structured interviews as a method is explained below.

The in-service teachers

The students were 45 in-service teachers of English from two separate semester-long (17-week) language courses that focus on grammar, pronunciation and vocabulary. The teachers took the courses because they were on a further education programme for in-service English teachers, most of whom had been teaching English for a number of years but had little (or no) formal qualification in English. A possible limitation of this study is that all participant teachers were in an ongoing programme, so no data was gathered from in-service teachers who were formally qualified in English. On the other hand, corpus instruction is not normally part of teacher education, so even formally qualified teachers are not necessarily educated in it.

One language course was for 25 primary and lower secondary school teachers (years 5-10); they will be referred to as “the 5-10 group”. The other course was for 20 secondary school teachers (years 8-13); they will be referred to as “the 8-13 group”. Teachers of these years were chosen because they teach English above beginner level, and the higher the year, the more a teacher tends to use digital tools (Gilje et al., 2016, p. 52). During the course, the students were asked if they wished to volunteer for an interview about corpora. The four who volunteered were subsequently interviewed and their contributions anonymized using assigned names from the random name generator at behindthename.com/random: Ebba, Rebekka, Amanda, and Katerina. The first three were from the 5-10 group, Katerina from 8-13. At the time of the interview, Ebba was teaching English to years 4 and 6, and had four and a half years' experience of teaching English in the Norwegian school system; Rebekka was teaching year 8 (previously having taught year 10), and had six years' experience; Amanda was teaching year 6 and had less than a year's experience; and Katerina was teaching vocational English in the first year of upper secondary school, with “three or four” years' experience.

An advantage of in-service informants is that their perspectives come from a position closer to the school classroom than that of pre-service informants. A disadvantage is that the interviews occur in the context of corpus instruction, so there is a hypothetical aspect to how they perceive corpora could be used in their teaching, in the same way that pre-service teachers' intentions for corpus use may not work out in practice (Farr, 2008, p. 39; Zareva, 2017, p. 75). However, in-service teachers are the agents for bringing corpora into a school teaching context, and corpus linguists have recognized that in-service teachers best understand pedagogical needs (Braun, 2007, p. 326; Römer, 2009, p. 83). In-service teachers have experience of the classroom and class preparation, and an informed judgement of how activities fit with curricula and lessons.

The corpus seminars

There were two seminars devoted to corpus instruction per language course, the first scheduled at the start. It was not institutionally possible to have more seminars, or corpora in more courses, or a dedicated corpus course. Pre-service corpus instruction in previous research involved more teaching (11 sessions in Breyer, 2011; 7 weeks' online teaching in Ebrahimi & Faghieh, 2016; 13-15 90-minute classes in Leńko-Szymańska, 2014; 30 75-minute classes in Zareva, 2017). As time can be linked to building confidence in the corpus user (Leńko-Szymańska, 2014, pp. 271-272), teachers' skills were expected to be limited at the time of the interviews. Nevertheless, this does not indicate that their knowledge about corpora is insufficient for them to be able to present in-service perspectives. Language classroom experience in schools has not yet contributed enormously to understanding of what is useful and what is challenging about corpus use, and informants can be considered able to relate what they learn to their practice, even after limited corpus instruction.

Each course had its first seminar, "Vocabulary and the Use of Language Corpora", in September 2018 (75 minutes for 5-10; 90 minutes for 8-13), and its second, "Grammar and the Use of Language Corpora", in October 2018 (3 hours for each group). In between the seminar dates, the 8-13 group had a corpus task as its obligatory vocabulary assignment, which kept them using corpora in the interim (it was not possible to add an obligatory task to the structure of 5-10). For both groups, written guides to the software were created (with screenshots), and instructional walkthrough videos were filmed. The guides and videos were to compensate for the low number of seminars, and also to help the students remember how to use software without its functions becoming difficult to remember. The corpus seminars were created to fit course plans, allowing corpora to be part of vocabulary and grammar teaching. It is not unusual for corpus methods to be attached to courses on grammar (Heather & Helt, 2012; Zareva, 2017). It seemed appropriate to also use corpora with vocabulary teaching because this was perceived useful in some of the previous research (see above) and work with corpora has also led to measurable vocabulary learning (Cobb, 1997). However, ways of using corpora in teaching have "usually focussed on specific aspects of language learning, e.g. vocabulary acquisition or specific aspects thereof", which leaves communicative aspects of language learning aside (Braun, 2005, p. 52), thus leaving a focus on vocabulary open to question.

Five corpus tools were introduced in the seminars. These were:

- The *Open American National Corpus (OANC)*, a 15 million-word "collection of American English, including texts of all genres and transcripts of spoken data produced from 1990 onward" (Anc.org). A text version and a tagged version were prepared for the students. The purpose of using a freely downloadable corpus was for teaching the use of a concordancer.
- *LancsLex*, an online tool for vocabulary teaching (Brezina & Gablasova, 2015). A text is pasted in, then it identifies the presence of the 2,500 most frequent English words in the text.
- *AntConc* (Anthony, 2018), a downloadable concordancer. It was decided to teach the use of a concordancer because in a Norwegian survey of in-service English teachers (Kavanagh, 2021, p. 11), 157 out of 188 informants claimed to collect pupil texts to get an overview of pupil mistakes, therefore some interest in do-it-yourself (DIY) corpora (Millar & Lehtinen, 2008) was anticipated, in which case a concordancer for searching a teacher's own corpus of texts might be welcome. The

choice of AntConc and the OANC was discussed with one of the compilers of corpus-analysis.com, who test and recommend corpus linguistics software.

AntConc was also used in courses in previous studies (Breyer, 2011, p. 158; Ebrahimi & Faghieh, 2016, p. 123; Leńko-Szymańska, 2014, p. 265). Reservations about using AntConc were: it might not be considered user-friendly if usability criteria recommended for software engineering were applied (Nielsen, 1993), nor was it designed to meet these criteria; and it involves other software (AntFileConverter; TagAnt), so there are multiple procedures to learn.

- *AntFileConverter* (Anthony, 2017) converts pdf and docx files into plain text for use with AntConc.
- *TagAnt* (Anthony, 2015). While an untagged corpus in AntConc could be used for vocabulary and some grammar work, TagAnt was required for tagging a corpus for most grammar work.

The first seminar consisted of an introduction to vocabulary, work with LanksLex, and work with simple searches of the OANC using AntConc. With 5-10, there was also time to work with case sensitive searches.⁵

The second seminar consisted of introductory slides, work with AntFileConverter, a reflection on the vocabulary assignment (8-13 only), work with copying concordance lines from AntConc, a grammar task with the untagged OANC, work with TagAnt, and a grammar task with the tagged OANC.

The interviews

The data consist of semi-structured interviews with four in-service teachers. Each was interviewed separately, between December 2018 and March 2019. Each informant was given an information letter about their rights of participation, the use of the data, and signed a consent form; the Norwegian Centre for Research Data (NSD) approved the interviews.

Corpus courses featured in previous research issued post-course questionnaires to informants (Breyer, 2011, pp. 185-205; Leńko-Szymańska, 2014; Zareva, 2017). This study had the opportunity to collect more elaborate data than that, with more detail, through semi-structured interviews. This type of interview makes several phenomena possible: a relationship between the researcher and informant; the back-and-forth of conversation; the chance of eliciting tacit understanding that is not stated directly; the chance of exploring the unexpected when conversation is not predetermined; an active role for informants; and the use by informants of their own words (Borg, 2015, p. 237). A particular contextual advantage of these interviews was that corpus software and interfaces could be used in the interview setting.

One informant (Ebba) was interviewed through a screen (using Skype for Business); the others had face-to-face interviews on campus. All interviews were in English. As English teachers, the informants have high proficiency in the language, in a country that generally has high proficiency (Education First, 2021). English was the working language of the course, which was the contextual backdrop of the interviews, and the basis of the interviewer-interviewee relationship. The terminology of language teaching and corpus linguistics were most familiar to the informants in English.

⁵ For example, if a user can search for *But* as distinct from *but*, data can be collected on sentences that begin with that conjunction.

The duration of each interview was approximately one hour. Recordings were transcribed. For each interview, a guide was created, consisting of approximately 25 questions. A composite of the four interview guides is the Appendix to this article. The interviews were conducted in the manner of semi-structured interviews described above, even though guide questions were asked.

When discussing challenges to the use of corpora, informants were asked “Are there computer-related challenges to using corpus methods, in your case or in the case of your pupils?”, ensuring there was some discussion of computer and IT challenges in this context. Otherwise, informants were not prompted towards answers to the research question. They were asked generally about authentic texts in ELT, and what technology they and their pupils have access to, as can be seen from the interview guides. The guides also show that informants were asked about their own knowledge of corpus methods; about whether informants collect pupil texts; about textbooks; about digital tools; about the tools introduced in the seminars; and about “competence aims” quoted from the subject curriculum (Utdanningsdirektorat, 2013).⁶ The informants were asked about the latter to see whether they would link aims to corpora.

During the interviews, the informants were shown the online corpus interfaces Sketch Engine for Language Learning (SKELL) and Netspeak (<https://netspeak.org/>). SKELL is a billion-word text corpus that consists of “sentences sorted according to their text quality”, from selected corpora and websites (Baisa & Suchomel, 2014). Netspeak, created at Bauhaus-University Weimar, searches Google Books. In the three on-campus interviews, informants were also shown the Corpus of Contemporary American English (COCA) (Davies, 2008-), which has been used in courses in previous research (Ebrahimi & Faghieh, 2016, p. 123; Leńko-Szymańska, 2014, p. 266; Zareva, 2017, p. 72). SKELL, Netspeak and COCA were included in the interviews because each was used by at least one corpus-using in-service teacher in another study (Kavanagh, 2021, p. 13), a finding not known to the researcher prior to designing the language course seminars for this study, but the students on the course were provided links to Netspeak, SKELL and COCA later in the semester. The informants had not investigated Netspeak, SKELL and COCA by the time of the interviews, but tested out the interfaces during the interviews.

The interviews were analyzed manually and the informants' answers were connected to the categories described above. These categories are based on the previous studies reviewed in this article and are thus not created specifically for the analysis of the interviews. However, they offer a link between the data discussed in this article and previous studies conducted on different populations.

Results

The results show that informants spoke about both the usefulness and challenges of using corpora. Corpus knowledge was limited: Ebba, the years 4 and 6 teacher, said “I wouldn’t say I know that much after just the seminars we had, but... it is good to know that it exists”; Rebekka (year 8) said she knew what she had been taught; Amanda (year 6) agreed she “can do a search”; and Katerina (first year secondary vocational school) said her “own competence isn’t great”. It seems the seminars were not enough; while basics were understood, informants lacked confidence. This is comparable to a pre-service course described in previous research (Leńko-Szymańska, 2014, pp. 271-272). Nevertheless, informants’ views are valuable because of their in-service experience.

⁶ Since superseded by Utdanningsdirektoratet, 2019.

Usefulness of corpora in teaching

Usefulness was categorized into *teachers' language awareness, vocabulary, authenticity and benefits of learner-corpus interaction*. Informants discussed teachers' language awareness, but interrelatedly with vocabulary: Ebba and Rebekka perceived benefits to teachers' language awareness only in the context of finding whether words collocate (Ebba: "You, as a teacher, have questions yourself"). Rebekka connected using corpora for this with lesson planning, and both she and Ebba thought corpora could prove the correct collocation of prepositions to pupils: for example, Rebekka spoke of year 10 pupils making collocation mistakes, "usually prepositions", and when told they were incorrect, the pupils disbelieved her, wanting proof. However, Rebekka has been able to assuage doubting pupils with examples from a dictionary, so was less inclined towards corpora, as "I don't know [if] some of the students would maybe have liked to see... 8,000 hits." The seminars covered how corpora differ from dictionaries, and they seemed to understand this. "You don't get every single use of a word in a dictionary," said Ebba. Nevertheless, informants showed a tendency to consult dictionaries first.

Amanda, Ebba and Rebekka said they would like to use SKELL and Netspeak, though Amanda emphasized a similarity between SKELL and dictionaries: "You can see the sentences, you can [ask] 'Okay, should I use this word? Should I not use this word?' You can go and see there. But for this I also use Macmillan or other online dictionaries." Amanda was able to see an advantage of SKELL over dictionaries, however. When asked what she would use to find many examples of a word, she indicated SKELL and added that she would not need a dictionary definition of the word alongside it, because she would "get the meaning quite easily" from SKELL's examples.

Amanda and Katerina saw the benefits of corpora for word frequencies, or as Katerina put it, "how common it [a word] is, or if it's completely obscure and it's a word you should not really be using". Amanda was the only informant who praised the advantages of LancsLex: "Checking out texts, I could have used it, to find the right text, if I don't use the textbook, then I could use it if I want to be sure... [For use with pupils] I still have to make a judgement whether or not it's suitable for that age group, 'cause this one doesn't help me with [the] age group, but it helps me to see if the words are of a high frequency."

While all four informants saw the benefits of corpus work with vocabulary, they were not unanimous about work with other language elements, notably grammar. Grammar teaching and learning was not one of the categories of usefulness that emerged from the previous research, but it was the focus of half of the corpus teaching. Only Amanda used the tagged OANC for work with grammar (for her own grammar learning on the language course). Rebekka saw a use for Netspeak with grammar, namely its Word Order feature: "A lot of my... students put the verb before the subject."⁷ Rebekka considered which pupils could use it: "...some of my [year 8] students could try it now. I think it's about maturity more than age, maybe. Some of them are quite eager ... they're curious language learners." Thus, it is not necessarily the language proficiency level of the year 8 pupils which determines the potential uptake from the use of Netspeak, but pupils' individual interest in language.

⁷ In Norwegian, the first language (L1) of most of Rebekka's pupils, the verb precedes the subject in main declarative clauses with fronted elements. Norwegian is a V2 language, which differs in word order to English, an SVO language.

Informants touched upon benefits of learner-corpus interaction, but only when considering Netspeak for pupils. In Netspeak, a question mark is used as a symbol for representing any letter or number (a “wild card”), which is an asterisk in AntConc.⁸ “That one’s easier,” said Rebekka about the question mark, and Amanda was enthusiastic about Netspeak for year 6 specifically because of this wild card: her pupils ask her “all the time” what word they can use with another.

Informants did not discuss authenticity in relation to corpora. When asked “What do you think of the use of authentic English texts in English language teaching?”, they said it was “necessary” (Katerina), “probably a good idea” (Rebekka), and less “boring” than the textbooks pupils “hate” (Amanda), but Amanda noted it was difficult to find texts at the appropriate level for her pupils. Ebba considered literature for authentic English. None of the informants mentioned corpora as a possible source of authenticity.

In the seminars, the students had been shown how to create their own corpora, with the purpose of, for example, finding common learner mistakes. Ebba, Rebekka and Amanda all said that, from the seminars, they saw the usefulness of finding common mistakes in a collection of pupil texts using AntConc, although of the four, only Ebba and Rebekka collected pupil texts. That they could see the point of this activity is comparable to Leńko-Szymańska’s pre-service informants enjoying creating their own corpora (Leńko-Szymańska, 2014, p. 271), although Rebekka found the process cumbersome (see below). Katerina agreed it was an advantage that one can create corpora for AntConc, but she did not see the use of a DIY corpus of pupil texts. Nevertheless, she thought that using AntConc with a corpus had advantages over COCA. Although she understood that COCA online is more up-to-date than the downloaded OANC, she liked knowing the location of the source texts for the latter.

The informants indicated that corpora were most useful for vocabulary teaching and learning. This may reflect that the first seminar focused on vocabulary, but the informants did not see as much usefulness for grammar teaching and learning, despite that being the focus of the second seminar. Beyond vocabulary, the informants’ perspectives do not substantially relate to the other categories of usefulness. This could be due to a lack of knowledge of corpora, but it could indicate a lack of relevance of some uses to in-service teachers. Additionally, even with what in-service teachers did find useful, limitations were noted: corpora would be useful only if a dictionary was not, or if the pupils would be interested in the information.

Challenges of using corpora in teaching

Challenges were categorized into *usability*, *computer and IT challenges*, *learner-corpus interaction challenges*, *lack of teacher need* and *workload*. The informants’ perceptions are related to four of these. They did not raise *workload* as an issue; Rebekka and Amanda did speak of time as a challenge (see below), but not in terms of preparing teaching materials or time in the classroom.

Rebekka thought using AntConc (AntFileConverter and TagAnt included) was “a bit *tungvint* [cumbersome]. It wasn’t that easy to do, there were a lot of processes, you had to convert and then import [files]”. For a DIY corpus, “since all our texts are in *It’s Learning* [a learning platform]... I’d first have to download them and then upload, it’s... a lot of processes.” She thought AntConc was time-consuming, as did Amanda. This is a usability issue because the criticism was meant for AntConc specifically. Amanda was more impressed

⁸ Wild cards were mentioned as one of the “technical difficulties” of corpora by Zareva’s informants (Zareva, 2017, p. 76).

with COCA, for being online (“I think this could be a more useful tool for me”). Amanda thought corpus tools were not as intuitive to use as Google or Microsoft products, but she did praise SKELL, which shows sentences rather than concordance lines. She said, “Visually, this appeals to me... I can see full sentences, yes... When you do a search it’s closer to a Google search.”

Rebekka also commented, “These programs, they’re not very intuitive,” and called AntConc “kind of confusing”, yet added optimistically that a program can be adequate “once you get it”. Rebekka thought results of SKELL searches were “a lot tidier” than AntConc or COCA, and “a lot easier to use”.

Katerina described AntConc searches as “a bit complex... probably I’d go to the dictionaries first. I go to *Oxford* or *Cambridge* or online dictionaries.” When discussing corpus interfaces generally, she compared them unfavourably with web browsers: “The way to get around to start using it is a bit harder than to just put the word in your [web browser]... so I think that’s probably one of the disadvantages, it’s not that self-explanatory, it’s not that user-friendly before you get your head around how to use it.”

All informants and their pupils have individual access to computers and IT resources, and have an internet connection. It was not infrastructure but IT skills that were perceived as a challenge. Ebba thought AntConc “easy” for herself to use, but saw a technological challenge for other teachers: “Many of my colleagues have still a lot of challenges when it comes to use of computers and programs... just the fact that it is a digital tool.” Rebekka also referred to “people who don’t enjoy using computers”. This is comparable to the “lack of IT skills”, and teachers uninterested in new methods, mentioned by Breyer (2011, p. 150) in a German context. Computer-related problems, which Breyer’s pre-service teachers had on their corpus linguistics course (Breyer 2011, p. 207), also occurred in the present study’s seminars: the groups had problems with downloading, with unzipping files, and with laptop power cables.

None of the informants mentioned, or said they considered, corpora for fulfilling “digital tools” and “digital resources” competence aims in the subject curriculum (see Appendix). They seemed to fulfil these aims in other ways, by using textbooks’ online resources, Quizlet (“to learn new words and conjugate verbs” – Ebba), Google Translate, G Suite for Education or Oxford Owl.

The informants teach at different stages in the education system and the needs of pupils at different stages of ELT, both within the same class and across different years, present obstacles to pupil-corpus interaction. Ebba thought that both LancsLex and AntConc were “really good”, but “maybe not for the level that I’m teaching”. She said that in years 5-7, pupils learn “basic grammar”, which could make corpora useful, but “they are not mature enough to freely give them access to a computer during a class”. AntConc itself presents a further problem, as for Ebba “it would be a challenge to install all these programs on their computers”. Rebekka said, “AntConc is for the teacher and not the students, probably.” Katerina does not see AntConc as appropriate even for her upper secondary pupils, because the usability issues of reading concordance lines: “It’s not self-explanatory... half a sentence at the front, and half a sentence at the back... the word that you have [searched for] is blue. What does the red mean? What does the green mean? ... For the level of pupils that I’m teaching, I think this is more at maybe university level.”

Ebba thought SKELL and Netspeak would be easier to use with pupils, because they involve merely opening a web page and typing. Rebekka thought SKELL’s Examples feature would be good for explaining the usage of words or expressions, but she meant this for pupils

at “the higher levels”. She explained it would be too much for her year 8 pupils, because when they search for a word in an online dictionary like Ordnett, the pupils ignore examples of usage and just read the first given meaning. If the pupils have trouble using the dictionary as instructed, getting them to read SKELL examples might be difficult. SKELL’s Word Sketch feature Rebekka ruled out entirely as unhelpful, Katerina thought pupils would not understand grammatical terminology used in the Word Sketch (while Amanda thought her year 6 pupils would at least understand “noun” and “verb” in it.) Katerina thought that even though a word can be found more easily in SKELL than in an online dictionary, the latter had an advantage over SKELL: she found example sentences in SKELL, even for simple searches like *dog*, to be context-free and not understandable. Katerina also did not like the way it was not clear why example sentences in SKELL appear in the order they do (she wondered why a search for *oracle* would produce sentences beginning *oracle implementation* as the third, eleventh and twelfth examples). Katerina would not even use SKELL for herself, preferring AntConc.

Despite thinking year 8 pupils could use Netspeak, Rebekka thought some of her pupils would not: “Students who really need a lot of help, they can’t be bothered... That’s basically the main problem. But I think maybe ninth grade or tenth grade, some students could use it, but I don’t think all of them would be interested.” Even when there is a tool that seems easy for pupils to use, pupils may not possess curiosity about language.

Amanda did not see the need to use corpora in year 6, “because they haven’t started writing long texts yet... If I were in grade 8 to 10, then I can see the benefits increasing. I think we have too short texts, it’s quite easy for me to [see what is in them].” Rebekka could not see the use when pupils are below year 10: “Common mistakes are quite obvious in year 8, and then for year 10 they write longer texts or more complex texts and it could be more interesting to look at the patterns in the texts.” This is reminiscent of informants in previous research who stated they would not need corpora in contexts when they were already familiar with common pupil mistakes (Kavanagh, 2021, p. 18).

To sum up, the challenges seen by these in-service teachers matched the challenges noted in previous research, in four categories: *usability*, *computer and IT challenges*, *learner-corpora interaction challenges* and *lack of teacher need*. The perceived challenges do not seem to differ much whether the teacher is a teacher educator, pre-service teacher, higher education teacher or in-service teacher. A view that seems particular to in-service teachers, categorized as *lack of teacher need*, is that informants would only use corpora when they see it as pedagogically useful.

Discussion and concluding remarks

Informants found corpora useful for teaching and learning vocabulary. In previous studies, corpora’s usefulness for vocabulary was a view held by both pre-service and in-service teachers. While that may indicate inherent usefulness of corpora with vocabulary teaching and learning, Leńko-Szymańska voiced concern about pre-service teachers addressing mainly vocabulary, and not other aspects of language: “it was impossible to establish if this focus on lexis and phraseology was a result of their lack of confidence in analysing corpus data for other features” (Leńko-Szymańska, 2017, p. 233). With the present informants, it is equally difficult to know whether inexperience leads them to see vocabulary teaching and learning as the main use of corpora. From what we know of corpus-using in-service teachers from previous research, they have found corpora useful also for grammar, as well as for checking acceptability of usage and for varieties of English (e.g. variation in

spelling) (Callies, 2019, pp. 250-252; Kavanagh, 2021, pp. 13-16), which is a wider range of uses, perhaps indicating that that experience with corpora increases the number of uses that in-service teachers will find.

What teachers consider challenging about corpora may indicate why corpora are not directly applied widely. Informants perceived most of the same challenges as found in previous research: *usability*, *computer and IT challenges*, *learner-corpus interaction challenges* and *lack of teacher need*. In-service teachers thus seem to see the same obstacles to corpus use as other types of teachers. This may indicate that these challenges are inherent to corpus linguistics methods and tools. To put a positive spin on that: if the same problems exist for every type of teacher, successful solutions may apply to all.

One suggested solution is “creating corpora that are pedagogically motivated, in both design and content, to meet pedagogical needs and curricular requirements so that corpus-based learning activities become an integral part, rather than an additional option, of the overall language curriculum” (McEnery & Xiao, 2010, pp. 374-375). Pedagogic corpora would sidestep challenges related to the use of general corpora in teaching, especially learner-corpus interaction challenges, but successfully relevant pedagogic corpora for Norwegian schools do not yet seem to exist. A study involving a set of pedagogic corpora (BACKBONE) in the secondary school classroom in Norway showed a need for an understanding of “what teachers do in the classroom”; the corpora struggled to be interesting and relevant to pupils (Karlsen, in preparation), especially at the stage when the English subject becomes more topic-focused than language-focused. Curricula in other countries may be more compatible with a focus on linguistic data (see Braun, 2007, p. 310; Pérez-Paredes, 2020, p. 75; and in an L1 context, Sealey & Thompson, 2007), but the aforementioned study also questions “student language interest and curiosity... as a motivational drive” (Karlsen, in preparation), and elsewhere it has been argued that “learners cannot be expected to be captivated by analysing corpus data” (Frankenberg-Garcia, 2014, p. 12), mirrored in the present study where Rebekka differentiates between pupils who are and are not interested in language.

Another suggested solution is the creation of a concordancer for the classroom (Breyer, 2011, pp. 220-223). The informants in the present study particularly focused on usability challenges in relation to AntConc, a concordancer, so this seems appropriate. No such software is currently available. Were it to exist, overcoming the usability challenges would be a welcome, but incomplete, solution. The corpus that is used with the concordancer must still be interesting and relevant to pupils.

Attention needs to be paid not only to usability, but the skills of the user. The present study’s informants perceived that there were both teachers and pupils without adequate IT skills for using corpora. Research indicates that one can expect digital skills to vary among teachers (Røkenes & Krumsvik, 2016) and pupils (Ståhl, 2017). Corpora could flourish better in a context in which digital skills in general were improving.

The final category of challenges, *lack of teacher need*, seems to imply limits to the pedagogical usefulness of corpora. But even if the kind of linguistic data that emerges from corpora is not relevant at all times, there are occasions when it is required. Since this study was undertaken, a new English subject curriculum came into effect, stating, “Language learning refers to developing language awareness and knowledge of English as a system...” (Utdanningsdirektoratet, 2019, p. 2). One could ask how teachers are expected to achieve this. It seems an opportunity for corpora to become especially useful. It would be fruitful to discover how corpora, even if that means only basic use of Netspeak and SKELL, can help.

The informants can see the usefulness of corpora, even if only for the teaching and learning of vocabulary, and this affirms that corpora can have a role in teaching. Furthermore, the informants of the present study are not necessarily representative of in-service teachers in general and they are too few to draw any conclusions; further research into the perspective of in-service teachers on corpora may involve others who use or want to use corpora for teaching grammar, varieties, and so on. A wider study of in-service teachers' engagement with corpora might reveal what is most appropriate to fit their needs and this should inform teacher education.

Further research could involve not only more in-service teachers, but could take place over a longer period of time. It is a concern that the students on the language course were confronted with much technology new to them, and in a programme in which it was not possible to have more seminars, to incorporate corpora into more courses, or have a dedicated corpus course. Judging by the informants' lack of confidence with corpora, the seminars were not enough. Over a longer period, their skills would develop more.

What these informants were required to learn could also be altered. The present study's informants were comparatively positive about Netspeak and SKELL, which unfortunately they were not introduced to during the language course seminars. An improvement in such seminars could be to focus on Netspeak and SKELL rather than other examples of corpora. This would better fulfil the recommendation of a previous study (Kavanagh, 2021, p. 20) to feed back what already works for some teachers into corpus instruction, as well as answer a call to learn from "ordinary teachers" (Chambers, 2019, p. 472) in a corpora and language learning context.

References

- Anc.org. The Open American National Corpus. <http://www.anc.org/>. Contents described: <http://www.anc.org/data/oanc/contents/>.
- Anthony, L. (2015). TagAnt (Version 1.2.0) [Windows]. Waseda University. <https://www.laurenceanthony.net/software>.
- Anthony, L. (2017). AntFileConverter (Version 1.2.1) [Windows]. Waseda University. <https://www.laurenceanthony.net/software>.
- Anthony, L. (2018). AntConc (Version 3.5.7) [Windows]. Waseda University. <https://www.laurenceanthony.net/software>.
- Baisa, V., & Suchomel, V. (2014). SKELL – Web Interface for English Language Learning. In *Eighth Workshop on Recent Advances in Slavonic Natural Language Processing*, 63-70. Brno: Tribun EU. Details of the English Corpus: <https://www.sketchengine.eu/english-skell-corpus/>.
- Borg, S. (2015). *Teacher cognition and language education: Research and practice*. Bloomsbury.
- Boulton, A., & Cobb, T. (2017). Corpus use in language learning: A meta-analysis. *Language Learning*, 67(2), 348-393.
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47-64.
- Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19(3), 307-328. doi:10.1017/S0958344007000535.

- Brezina, V., & Gablasova, D. (2015). English vocabulary tool. <http://corpora.lancs.ac.uk/vocab>.
- Breyer, Y. (2011). *Corpora in Language Teaching and Learning: Potential, Evaluation, Challenges*. Peter Lang GmbH, Internationaler Verlag der Wissenschaften. https://www.researchgate.net/publication/264499058_Corpora_in_Language_Teaching_and_Learning_Potential_Evaluation_Challenges.
- Callies, M. (2019). Integrating corpus literacy into language teacher education: the case of learner corpora. In S. Götz, & J. Mukherjee (Eds.), *Learner corpora and language teaching* (pp. 245-263). John Benjamins Publishing Company.
- Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Language Teaching*, 52, 460–475. doi:10.1017/S0261444819000089
- Cobb, T. (1997). Is there any measurable learning from hands-on concordancing? *System*, 25(3), 301-315.
- Crosthwaite, P. (2020). *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. Routledge.
- Davies, M. (2008-). The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. <https://www.english-corpora.org/coca/>.
- Ebrahimi, A., & Faghih, E. (2016). Integrating corpus linguistics into online language teacher education programs. *ReCALL*, 29(1), 120–135. doi:10.1017/S0958344016000070
- Education First. (2021). *English Proficiency Index 2021*. <https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2021/ef-epi-2021-english.pdf>
- 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.
- Frankenberg-Garcia, A. (2014). How language learners can benefit from corpora or not. *Recherches en didactique des langues et des cultures*, 11(1), 1-15.
- Gilje, Ø., Ingulfsen, L., Dolonen, J. A., Furberg, A., Rasmussen, I., Kluge, A., Knain, E., Mørch, A., Naalsund, M., & Skarpaas, K. G. (2016). *Med ARK&APP – Bruk av læremidler og ressurser for læring på tvers av arbeidsformer*. University of Oslo.
- Heather, J., & Helt, M. (2012). Evaluating corpus literacy education for pre-service language teachers: six case studies. *Journal of Technology and Teacher Education*, 20(4), 415–440.
- Karlsen, P. H., & Monsen, M. (2020). Corpus Literacy and Applications in Norwegian Upper Secondary Schools: Teacher and Learner Perspectives. *Nordic Journal of English Studies*, 19(1), 118-148.
- Karlsen, P.H. (In preparation). Integrating Multimedia Corpora in the Secondary School Classroom in Norway.
- Kavanagh, B. (2021). Bridging the Gap from the Other Side: How Corpora are Used by English Teachers in Norwegian Schools. *Nordic Journal of English Studies*, 20(1), 1-35.
- Koellner, K., & Greenblatt, D. (2018). Inservice Teacher Education entry in Oxford Bibliographies. <https://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0196.xml>
- Leńko-Szymańska, A. (2014). Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *ReCALL*, 26(2), 260-278.
- Leńko-Szymańska, A. (2017). Education teachers in data-driven learning: Tackling the challenge. *Language Learning and Technology*, 2(3), 217-241.
- Lin, M. H., & Lee, J. (2015). Data-driven learning: changing the teaching of grammar in EFL classes. *ELT Journal*, 69(3), 264-274. doi:10.1093/elt/ccv010.

- Liu, D. & Lei, L. (2017). *Using Corpora for language learning and teaching*. TESOL International Association.
- McEnery, T. & Xiao, R. (2010). What corpora can offer in language teaching and learning. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning (Volume 2)* (pp. 364-380). Routledge.
- Millar, N., & Lehtinen, B. (2008). DIY learner corpora: Bridging gaps between theory and practice. *The JALT CALL Journal*, 4(2), 61-72.
- Mishan, F. (2005). *Designing Authenticity into Language Learning Materials*. Intellect Books.
- Morrow, K. (1977). Authentic texts and ESP. In S. Holden (Ed.), *English for Specific Purposes* (pp. 13-17). Modern English Publications.
- Nielsen, J. (1993). *Usability engineering*. AP Professional.
- Pérez-Paredes, P. (2020). The pedagogic advantage of teenage corpora for secondary school learners. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners* (pp. 67-87). Routledge.
- Røkenes, F. M., & Krumsvik, R. J. (2016). Prepared to teach ESL with ICT? A study of digital competence in Norwegian teacher education. *Computers & Education*, 97, 1-20.
- Römer, U. (2009). Corpus research and practice. What help do teachers need and what can we offer? In Aijmer, K. (Ed.), *Corpora and language teaching* (pp. 83-98). John Benjamins Publishing Company.
- Römer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225. doi:10.1017/S0267190511000055.
- Sealey, A., & Thompson, P. (2007). Corpus, concordance, classification: Young learners in the L1 classroom. *Language Awareness*, 16(3), 208–223.
- Ståhl, T. (2017). How ICT savvy are Digital Natives actually? *Nordic Journal of Digital Literacy*, 12(3), 89-108. doi: 10.18261/ISSN.1891-943X-2017-03-04
- Utdanningsdirektoret. (2013). English Subject Curriculum. <https://data.udir.no/k106/ENG1-03.pdf?lang=eng>.
- Utdanningsdirektoret. (2019). English Subject Curriculum. <https://www.udir.no/lk20/eng01-04?lang=eng>.
- Zareva, A. (2017). Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal*, 71(1), 69-79. doi:10.1093/elt/ccw045.

Appendix: Interview guide

Four informants: a year 4 and 6 teacher (“Ebba”), 2018; a year 8 teacher (“Rebekka”), 2019; a year 6 teacher (“Amanda”), 2019; and an upper secondary vocational school teacher (“Katerina”), 2019.

	Question	Asked of
1	In what year(s) do you teach English?	All four informants
2	For how long have you taught English in the Norwegian school system?	All four informants
3	How many pupils do you teach English to (per class)?	All four informants
4	What technology do the pupils have access to?	All four informants
5	What technology do you have access to?	All four informants
6	This was your first semester as a [further education] student. How was the experience?	All four informants
7	Do you ever base your teaching on common mistakes that your pupils make?	All four informants
8	How do you get an overview of the mistakes?	All four informants
9	If you collect pupil texts, how do you use them?	The informants who collected pupil texts were the year 4 and 6 teacher and the year 8 teacher
10	What do you think of the use of authentic English texts in English language teaching?	All four informants
11	What do you think of textbooks in English language teaching?	All four informants
12	What are your views on digital tools? Do you use them?	All four informants
13	Curricular competence aims after year 7. Language learning: enable pupils to “use digital resources and other aids in one’s own language learning”. How do you fulfil this aim?	Year 4 and 6 teacher Year 6 teacher
14	Curricular competence aims after year 7. Written communication: enable pupils to “use digital tools and other aids to find relevant information and to create different types of texts”. How do you fulfil this aim?	Year 4 and 6 teacher Year 6 teacher
15	Curricular competence aims after year 10. Language learning: enable pupils to “select different digital resources and other aids and use them in an independent manner in [their] own language learning”. How do you fulfil this aim?	Year 8 teacher
16	Curricular competence aims after year 10. Written communication: enable pupils to “use digital tools and formal requirements for information processing, text production and communication”. How do you fulfil this aim?	Year 8 teacher
17	Competence aims after years Vg1 and Vg2. Language learning: enable pupils to “evaluate	Upper secondary vocational school teacher

	different digital resources and other aids critically and independently, and use them in own language learning”. How do you fulfil this aim?	
18	Having been introduced to corpus methods, how much would you say you know about them?	All four informants
19	What different users do you think corpora could have?	Year 4 and 6 teacher
20	What do you think the advantages of corpus methods are?	All four informants
21	What do you think the disadvantages of corpus methods are?	All four informants
22	What do you think the challenges are to adopting corpus methods?	All four informants
23	Are there computer-related challenges to using corpus methods, in your case or in the case of your pupils?	All four informants
24	What did you think of LancsLex?	All four informants
25	What did you think of AntConc?	All four informants
26	What did you think of the Open American National Corpus?	All four informants
27	Did you look at the concordancer in Sketch Engine? If so, what did you think of it?	Year 4 and 6 teacher
28	What do you think of SKELL?	Year 8 teacher Year 6 teacher Upper secondary vocational school teacher
29	Did you look at Netspeak? If so, what did you think of it?	Year 4 and 6 teacher
30	What do you think of Netspeak?	Year 8 teacher Year 6 teacher Upper secondary vocational school teacher
31	What do you think of corpora as authentic language material?	Year 4 and 6 teacher
32	Did exposure to corpus methods make you think any differently than before? (If so, how? / If not, why not?)	Year 4 and 6 teacher
33	Here are some suggested reasons for why it might be useful to teach with corpora – do you agree or disagree? Self-discovery; different types of learners exist; corpora open up different channels/provide more input; corpora more visual; dictionaries don't have patterns; corpora have whole texts; corpora can show more recent language phenomena; corpora show variation is inherent; corpora as source of informal use; spoken corpora for conversation analysis.	Year 4 and 6 teacher
34	If there was a book of readymade language exercises for corpus methods, would you use it?	Year 4 and 6 teacher