Inland Norway
University of
Applied Sciences

Faculty of Education

Barry Kavanagh

How Corpora Can Be Useful in English Language Teaching in Norwegian Schools

PhD in Teaching and Teacher Education
2023

Barry Kavanagh

How Corpora Can Be Useful in English Language Teaching in Norwegian Schools

PhD Dissertation

2023

Faculty of Education

Inland Norway University of Applied Sciences

Printed by: Flisa Trykkeri A/S

Place of publication: Elverum

PhD Dissertation in Teaching and Teacher Education (PROFF) no. 33

ISBN printed version: 978-82-8380-457-7

ISBN digital version: 978-82-8380-458-4

ISSN printed version: 2464-4390

ISSN digital version: 2464-4404

Abstract

This article-based dissertation suggests how corpora – collections of texts of naturally occurring language, stored in an electronic database – can be useful in English Language Teaching in Norwegian schools. The focus of this research is on direct pedagogical corpus applications, meaning the interaction of teachers with corpora, pupils with corpora, or both. Previous researchers have found there to be a 'gap' between corpus linguistics and teaching practice. The long-standing metaphor that has been used for the resolution of this is 'bridging the gap'. This dissertation puts forward the idea that one can 'bridge the gap' by starting from the teachers' 'side' of the gap, by providing teachers with solutions for what is currently required in their English teaching: authentic language data, explicit language learning, language awareness, critical thinking, in-depth learning, and digital skills.

The first article presents data indicating there are few direct pedagogical corpus applications in Norwegian schools. The research involved a national survey of English teachers in primary and secondary school to discover how widespread the use of corpora is (34 out of 193 teachers answered they had done some work with corpora), and follow-up interviews obtained the perspectives of three corpus-using teachers: they used GloWbE, SKELL, Netspeak and COCA; teacher-corpus interaction was for reference and for preparing teaching (of vocabulary, and varieties of English); and pupil-corpus interaction was encouraged by two of the teachers.

The second article presents interview data from four teachers who had not used corpora, but were introduced to them through corpus seminars integrated by the present author into a language course for in-service teachers. The informants found corpora useful for teaching and learning vocabulary, and perceived the challenges to be: usability, lack of teacher IT skills, pupil-corpus interaction challenges (complexity of software and concordance lines; lack of pupil interest in language), and lack of teacher need (language mistakes being 'obvious' to teachers in the lower years of education).

The third article presents how corpora can be used in the English subject in lower secondary school in Norway, with the curriculum for English that came into effect in August 2020. Corpus exercises were designed by the present author, adapted from curriculum-relevant textbook exercises, and influenced both by how teachers have said they use corpora (see the first article) and by teacher perspectives on corpora (see the second article). A reason the corpus exercises

were adaptations of textbook exercises was to build upon what teachers already work with (there is a strong textbook focus in primary and lower secondary school levels in Norway). A second reason was to show that corpus-based teaching materials can replace, enhance or revise textbook exercises, in light of the curriculum (critical thinking, in-depth learning, knowledge of English as a system, and digital skills).

The dissertation concludes by discussing the types of suggested solutions there have been for 'bridging the gap' (more education of teachers, new corpora, new software, and collaboration between linguists and teachers). It adds its own suggested solution: a collection of bespoke corpus exercises matched to a specific curriculum and/or pupil level, which avoids challenging software or interfaces, uses free and accessible corpora, does not give the impression that the approach is only for linguists, and does not require prior teacher training to use it.

Keywords: corpora, corpus, corpus activities, corpus-based approach, corpus-based teaching materials, corpus exercises, corpus linguistics, curriculum renewal, data-driven learning, direct pedagogical corpus applications, English as a second/foreign language, English subject in Norway, English vocabulary teaching and learning, language learning, lower secondary school, materials development, Norwegian education, pre-tertiary education, textbook exercises

Sammendrag

Denne artikkelbaserte doktoravhandlingen utforsker hvordan korpus-samlinger av tekster av naturlig forekommende språk som er lagret i en elektronisk database, kan være en nyttig del av undervisningen i engelsk i norsk skole. Fokuset er på direkte pedagogiske korpusapplikasjoner, det vil si på læreres interaksjon med korpus, elevers interaksjon med korpus, eller begge deler. Tidligere forskere har funnet et 'gap' mellom korpuslingvistikk og undervisningspraksis. Metaforen som lenge har blitt brukt for å endre dette, er å 'bygge bro' over gapet. Denne avhandlingen fremmer ideen om at man kan bygge en slik bro ved å starte fra lærernes 'side' av gapet og gi lærere alternative tilnærminger til aktuelle utfordringer i engelskundervisningen, som bruk av autentiske språkdata, eksplisitt språklæring, språkbevissthet, kritisk tenkning, dybdelæring og digitale ferdigheter.

Den første artikkelen presenterer data som indikerer at det er få direkte pedagogiske korpusapplikasjoner i bruk i norsk skole. Det ble først gjennomført en nasjonal undersøkelse blant engelsklærere i grunnskolen og videregående for å finne ut hvor utbredt bruken av korpus er (34 av 193 lærere svarte at de hadde jobbet litt med korpus). Videre ble det gjennomført oppfølgingsintervjuer med tre korpusbrukende lærere. De brukt GloWbE, SKELL, Netspeak og COCA. Lærernes interaksjon med korpusene var å bruke dem for referanse og for å forberede undervisning (av vokabular og av varianter av engelsk). Elev-korpus-interaksjon ble oppmuntret av to av lærerne.

Den andre artikkelen presenterer intervjudata fra fire lærere som ikke hadde brukt korpus, men som ble introdusert for det gjennom egne seminarer som forskeren utviklet som del av et etterutdanningskurs for lærere. Informantene fant korpus nyttig for å undervise og lære vokabular. Av utfordringer med bruk av korpus oppfattet lærerne: manglende brukervennlighet, manglende IT-kompetanse hos lærere, interaksjonsutfordringer mellom elev og korpus (kompleksiteten til programvaren og konkordanslinjer; manglende elevinteresse for språk) og manglende behov hos lærerne (språkfeil er 'åpenbare' for lærere på de laveste utdanningstrinnene).

Den tredje artikkelen presenterer hvordan korpus kan brukes i engelskfaget i ungdomsskolen i Norge med utgangspunkt i læreplanen for engelsk som trådte i kraft i august 2020. Ulike korpusøvelser ble designet av forskeren. Disse ble tilpasset etter læreplanrelevante lærebokøvelser, både ut fra hvordan lærere svarer at de bruker korpus (se første artikkel) og ut fra

lærerperspektiver på korpus (se andre artikkel). En grunn til at korpusøvelsene var tilpasninger av øvelser fra lærebøker, var for å kunne bygge videre på det lærere allerede jobber med (det er et sterkt lærebokfokus i grunnskolen i Norge). En annen grunn var for å vise at korpusbasert undervisningsmateriell, i lys av læreplanen (kritisk tenkning, dybdelæring, kunnskap om engelsk som system og digitale ferdigheter), kan erstatte, forbedre eller revidere lærebokøvelser.

Avhandlingen avsluttes med å diskutere ulike typer foreslåtte løsninger for å bygge bro over det nevnte gapet (mer utdanning av lærere, nye korpus, ny programvare og samarbeid mellom lingvister og lærere). Den legger også til sin egen foreslåtte løsning: en samling skreddersydde korpusøvelser tilpasset en spesifikk læreplan og/eller elevnivå som unngår utfordrende programvare eller grensesnitt, bygger på gratis og tilgjengelige korpus, ikke gir inntrykk av at tilnærmingen kun er for lingvister og som ikke krever tidligere lærerutdanning for å brukes.

Nøkkelord: korpus, korpusaktiviteter, korpusbasert tilnærming, korpusbasert undervisningsmateriell, korpusøvelser, korpuslingvistikk, læreplanfornyelse, datadrevet læring, direkte pedagogiske korpusapplikasjoner, engelsk som andre-/fremmedspråk, engelskfaget i Norge, vokabularundervisning og -læring, språkopplæring, ungdomsskole, utvikling av materiell, norsk grunnutdanning, lærebokøvelser

Preface

My master's thesis, written in 2014, was a corpus investigation, and immediately afterwards I was employed in English teacher education. It was perhaps inevitable, then, that these two things would come together, and I would work on a Ph.D. dissertation that involved both corpus linguistics and English teachers.

This project was undertaken between 2017 and 2023. Now that the finished text finally exists, I declare here my thanks to my main supervisor, Professor Sylvi Rørvik, who did an incredible job of directing me towards the light at the end of the tunnel. She was particularly creative at suggesting what I could research when there were avenues that did not lead anywhere. I am also very grateful to the feedback and support of my second supervisor, Professor Marte Monsen. Additionally, I would like to thank Professor Lise Iversen Kulbrandstad, the study programme co-ordinator, who offered and gave me help several times.

During the process, Associate Professor Hildegunn Dirdal (University of Oslo) at the 50% evaluation, and Associate Professor Susan Erdmann (University of Agder) at the 90% evaluation, both gave crucial feedback, helping solidify the work and elucidating for me which parts were ready for completion. I am grateful to both of them.

From my life outside, my heart and thanks go to Helle Barstad Behrens for her love and support. It can't be easy to endure someone engaged in an inherently stressful research project. Also, owing to the pandemic and other catastrophes over the last six years, it has not always been possible to work on campus in Hamar, so personal gratitude also goes to these organizations: the University of Oslo, Stiftelsen Grünerløkka Lufthavn, and Lufthavna AS.

Barry Kavanagh

29 August 2023

Table of contents

List of tables

List of articles

Article 1

Kavanagh, B. (2021a), '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.

Article 2

Kavanagh, B. (2021b), 'Norwegian In-service Teachers' Perspectives on Language Corpora in Teaching English', *Nordic Journal of Language Teaching and Learning*, 9/2: 90-106.

Article 3

Kavanagh (in preparation), 'Corpus Exercises for Lower Secondary English'.

Chapter 1 Introduction

A corpus is 'a systematic compilation of naturally occurring language' (Friginal 2018: 12) and 'a collection of texts stored in an electronic database' (Baker, Hardie & McEnery 2006: 48). This dissertation concerns the use of corpora[1] in the English subject in Norwegian schools. Corpora have been the linguistic data used for dictionaries, for grammars, for defining a core vocabulary, for specialist vocabulary, and for exercise material for learners (Hasselgård 2020: 7-8), and these uses, in terms of pedagogical applications, have been called the 'indirect applications' of corpora, where researchers and materials writers interact with them (Römer 2011: 207). The focus of this dissertation is on 'direct applications', which involve the interaction of teachers and learners with corpora (ibid.). Some of the best-known corpora are 'general' or 'reference' corpora, 'designed to represent the language-at-large' (Friginal 2018: 16). A direct application might be a teacher using such a corpus to obtain data for a slide to show pupils the 20 most common words that go with the noun 'choice' (e.g. 'no' or 'your'), or a direct application might be a year 10 pupil exploring a corpus to understand adjective-noun collocations in English, for example, to find out whether it is more common to write 'heavy rain' or 'deep rain'. From these examples, it can be seen that a corpus shows what a dictionary or thesaurus does not. A dictionary or thesaurus will present the word 'rain', but will not show learners what other words 'rain' tends to occur (and not occur) in combination with. This dissertation explores how familiar teachers are with the pedagogical use of corpora, what the benefits are of this use of corpora, and what challenges teachers face in this regard (see the research questions below in 1.4).

The rest of this introduction is divided as follows. Section 1.1 clarifies first, the use of the term 'direct applications' in relation to another term used in this field, 'data-driven learning', and second, what direct applications mean in terms of dimensions of language learning. 1.2 and 1.3 focus on why corpora would be used in language learning: 1.2 covers the history of using corpora for this, and 1.3 describes the benefits of using corpora for it. However, despite these benefits, corpora do not seem to be directly applied often in schools. 1.4 explains why research on direct applications is needed, and introduces the research question and sub-questions. Finally, 1.5 outlines the structure of the overall dissertation.

[1] Singular *corpus,* plural *corpora.*

1.1 Direct pedagogical corpus applications

The term that has been used above, 'direct applications', more fully, 'direct pedagogical corpus applications', can be synonymous with the term 'data-driven learning' (DDL), but there exist definitions of DDL that specify inductive learning (Boulton 2011, cited in Crosthwaite 2020: xiv-xv), or learner-initiated research (Johns 1991: 1-3), while this dissertation has a broader focus, that is, on any use of corpora by pupils and/or teachers. There is concern about whether a definition of DDL that means pupils would be self-taught researchers can meaningfully apply to secondary school language learners (Braun 2007: 324), or younger learners, but if the broadest definitions of DDL are used, that it 'consists of using the tools and techniques of corpus linguistics for pedagogical purposes' (Gilquin & Granger 2022: 430), or that it means 'using the tools and techniques of corpus linguistics for second language learning or use' (Boulton & Cobb 2017: 348), then DDL is synonymous with direct applications.[2] There is good reason to use a broad term in this research project: any use of corpora by teachers or pupils, especially if it can potentially be developed or become widespread, is of interest here.

Direct applications of corpora can be either inductive learning or deductive learning. This is important to state, because pre-service teachers have mistakenly thought that using corpora in teaching is inductive learning only (Ebrahimi & Faghih 2016: 128). In the inductive approach, learners start with the corpus language data as evidence (Breyer 2011: 52), while in the deductive approach, learners begin 'with a previously learned rule' and go to the corpus for verification or practice (ibid.: 53; see inductive and deductive corpus activities in Liu & Lei 2017: 31-33 and Pinto et al. 2023).

Both inductive learning and deductive learning are part of explicit language learning (Dekeyser 2005: 242; Newby 2020: 222). Explicit language learning involves the learner's consciousness of the language structure being learnt. There are both explicit and implicit learning mechanisms in second language acquisition (Dekeyser 2005: 241). An implicit learning mechanism is that of a child learning its first language (ibid.), that is, a 'lack of consciousness of the structure being learned' (ibid.). A 'strong' version of communicative language teaching would see a second language being learnt that way too (Skulstad 2020: 56). The direct applications of corpora in language teaching are not implicit learning, they are explicit, whether the approach is inductive or

[2] While Römer's definition of 'direct applications' includes 'teacher-corpus interaction' (Römer 2011: 207), there are other scholars who exclude it, e.g. Yoon & Jo (2014: 97). This dissertation follows Römer in this respect. Also worth noting is that Römer defines direct applications as 'hands on for learners and teachers (data-driven learning)' (Römer 2011: 207). The equivalence of 'hands on' and 'data-driven learning' can be confusing here, because there exists 'hands off' interaction with paper-based concordances that is also considered to be data-driven learning (Vyatkina 2016).

deductive. In inductive learning, a pupil is given language examples, from which rules can discovered, while in deductive learning a pupil is given language rules, followed by exercises (Newby 2020: 222). In the case of corpora, direct applications involve exploring language data and consciously discovering a rule or phenomenon from that data, or exploring language data based on conscious, pre-existing knowledge of a rule or phenomenon. Consciousness makes direct applications explicit learning.

In summary, direct applications can be called DDL, depending on the definition. When used in practice, direct applications (teacher-corpus interaction, learner-corpus interaction) are explicit language learning, which can be either inductive or deductive. Some of the abovementioned concepts – communicative language teaching, explicit language learning, and differing learning methods – are important for understanding the English subject in Norwegian schools (see Chapter 2).

1.2 Corpora in language learning

Corpora have a long-standing association with language teaching and learning. Xu's historical overview of using corpora in English language teaching begins with John Freeman's corpus-based frequency list of 1820, created 'to teach adults to read' (Xu 2022: 12). Clearly, it seemed useful to rely on a corpus for evidence of what should be taught. Later came Ayres' 1913 work on the most frequent words used in correspondence, identifying 'the words that ordinary people need to know how to spell', in contrast to the USA's National Education Association's list of spelling words of the time, which relied on intuition rather than evidence (ibid.). Intuition can be unreliable because 'humans tend to notice unusual occurrences more than typical occurrences' (Biber, Conrad & Reppen 1998: 3). The quantitative corpus-based evidential approach continued with Thorndike's *Word Book* of 1921, Lorge's semantic frequency lists of the 1930s and 40s, and Fries's *American English Grammar* of 1940 (Xu 2022: 13-14). Then the computer corpus arrived with the Brown corpus in 1964. Xu recounts how corpora have been used for the development of reference books (pedagogical grammars, dictionaries) and course materials (ibid.: 14-18), all of which can be considered indirect applications. In terms of direct applications, this also has a long history. Leech 'began using an incomplete prototype LOB (Lancaster-Oslo/Bergen) Corpus for postgraduate teaching as early as 1976' (Leech 1997: 2). In the 1980s, Higgins and Johns's early work in Computer-assisted Language Learning 'availed itself' of corpora (ibid.: 3). In the early 1990s, Tribble and Jones 'started to experiment with printed concordances in language classrooms' (Xu 2022: 18), and Johns proposed the approach called Data-Driven Learning (DDL) (Johns 1991). Just as Ayres and Thorndike earlier in the century emphasized evidence

over intuition, 'the native English-speaking tutor (Johns himself) did not have the final say of grammatical correctness or acceptability, but the corpus evidence, especially collocational patterns, did' (Xu 2022: 19). In the early 1990s, learners used printouts of concordance lines in grammar and vocabulary learning. Later, access to computers became more possible. Resources have been 'upgraded' further with online corpus access (ibid.), which is how teachers and learners will mostly encounter corpora today.

This association of corpus linguistics with language teaching and learning has occurred because of advantages that corpus linguists have recognized. A quarter of a century ago, Leech identified these advantages to corpus tasks on computer: 'Automatic searching, sorting, scoring', 'Promoting a learner-centred approach', 'Open-ended supply of language data' and 'Enabling the learning process to be tailored' (Leech 1997: 10-11). More recently, Gilquin and Granger have listed these pedagogical advantages of corpora: they expose learners to authentic language data, they provide a point of comparison for learners' own writing, corpus searches include an element of discovery, and the exploration of language in this way develops learning skills and cognitive skills (Gilquin & Granger 2022: 2-3). Corpora have consistently been seen as beneficial for language learning. This leads to the question of how they can be beneficial, which is discussed in the next section.

1.3 The benefits of corpora in language learning

It is essential to know whether a corpus-based approach functions pedagogically. Boulton and Cobb undertook a meta-analysis, which 'synthesizes quantitative results in the form of effect sizes' (Boulton & Cobb 2017: 353), of 64 separate DDL studies. It was 'the first time that DDL work has been brought together for all to see and consider as a whole – including for DDL researchers' (ibid.: 388). The meta-analysis showed large effect sizes, 'which suggests that DDL does work' (Gilquin & Granger 2022: 14). However, it ought to be noted, because of the context of the school-level focus of this dissertation, that there was 'little research' available below tertiary education level to include in the meta-analysis (Boulton & Cobb 2017: 375). An empirical study by Braun based in secondary education was included (ibid.: 364). In that study, Braun argued for 'a move from data-driven learning to needs-driven corpora, activities and methodologies' (Braun 2007: 326). Recall (from 1.1 above) that there are differing meanings of 'DDL': here, Braun means a move away from DDL in the sense of learners as self-taught researchers, towards whatever the pedagogical need for corpora at secondary school level might be.

Thus, the learning context has importance in a discussion of benefits. This dissertation takes English language learning in Norwegian primary and secondary schools as the context. To focus

on some specifics of language learning at primary and secondary level, the benefits of direct applications would include: access to frequency information, information about conventional usage (meaning, collocation), practice in critical thinking, access to authentic language data, and, as a result of these, a greater degree of in-depth learning. Each of these concepts is elaborated on below.

Frequency information. A corpus enables learners to discover how often a word or phrase occurs, which helps them choose from alternative wordings. For example, if a learner was to choose between synonyms, a corpus could reveal how frequent or infrequent each synonym is, giving a sense of 'the safest option' (Hasselgård 2020: 5).

Information about conventional usage. To stay with the example of choosing between synonyms, a learner might be unsure which synonym to use. A corpus would provide data on how each synonym is used, and in what contexts it tends to appear. A corpus can supply much-needed examples of usage, and evidence from which the learner can work out nuanced meanings. Not only that, but a corpus could reveal collocations: what other words a chosen word or phrase tends to be used with.

Practice in critical thinking. Language data is retrieved from a corpus, and drawing conclusions from such data involves critical thinking. Even figuring out the software which searches a corpus can require critical thinking. From Norwegian educational documents that underlie the national curriculum (see Chapter 2), two points can be noted about critical thinking: first, that critical thinking is 'being able to reason and analyze, identify relevant questions, and being able to use relevant strategies for complex problem solving. It is also about being able to evaluate claims, arguments and evidence from different sources in complex and unfamiliar situations' (Norges Offentlige Utredninger 2015: 33, my translation); and second, that Norwegian education promotes the idea of developing learners' ability to think critically (Norwegian Ministry of Education and Research 2016: 18).

Access to authentic language data. An authentic text has been 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: 13, cited in Gilmore 2007: 98). In language teaching, there has been an argument for the use of authentic language rather than contrived textbook language. The problem with textbook language is that it can seem artificial (Gilmore 2004: 363). This can have practical consequences for learners. For example, direction-giving in natural conversations is more complicated than the 'standard, three-step, model presented to students (request for directions – direction-giving – thanks)' (Gilmore 2007: 102). Authentic texts have therefore

become valued (ibid.: 97). Corpora are seen as sources of authentic language (e.g. Römer 2011: 210), indeed they have 'fuelled interest' in authenticity in English language teaching (Buendgens-Kosten 2014: 457).

In-depth learning.[3] This can be 'defined as students' understanding of concepts and the relations between them, students relating new ideas to familiar concepts and principles in order for the new understanding to be used in problem-solving in new and unfamiliar situations' (Burner 2020: 55). Research indicates it is important to pupil development in learning (National Research Council 2000; Pellegrino & Hilton 2012; Sawyer 2006; all cited in Norwegian Ministry of Education and Research 2016: 33). In practice, in-depth learning can mean many things, one of which is that pupils can choose to go in-depth in a subject (Norges Offentlige Utredninger 2015: 11), with the school providing sufficient time for specialization, as well as support and guidance (ibid.). Arguably, all the other benefits described above can collectively feed a pupil's in-depth learning.

In addition to being beneficial for language learning, direct applications are warranted by the *Core Curriculum – values and principles for primary and secondary education* (Norwegian Directorate for Education and Training[4] 2017), and the English subject curriculum (Norwegian Directorate for Education and Training 2019), where some of the above concepts apply (authenticity, critical thinking, and in-depth learning). The curricular context will be described in Chapter 2.

1.4 The research, the research question, and sub-questions

The compelling reason to research direct applications is that corpora do not seem to be directly applied very often by teachers in schools (Callies 2019). Internationally, researchers have found there to be a 'gap' between corpus linguistics and teaching practice, and have discussed the difficulty of 'bridging the gap' (Mukherjee 2004; Breyer 2011: 146). Prior to the present research project, it has not been established how much of a 'gap' there is between corpus linguistics and teaching practice in the Norwegian school context. Internationally, suggestions for 'bridging the gap' have included educating teachers, compiling pedagogically motivated corpora, and improving usability (see Chapter 3). While these are good suggestions, they have not so far resulted in increased corpus use. There has been a suggestion that linguists promoting the use of corpora in language teaching could focus more on educational principles and curricular context (Meunier 2022: 348). Researching direct applications within the specific educational and curricular context of Norway helps to develop this focus. Those wishing to spread the use of

[3] Sometimes translated from Norwegian *dybdelæring* as 'deep learning'.
[4] The Directorate is the executive agency for the Norwegian Ministry of Education and Research.

corpora need to explore the English subject classroom. The focus of this dissertation is the use of corpora from the perspective of teaching practice. This means that the perspective of in-service teachers[5] is important, because of their centrality to the language classroom. A fundamental question for teachers is whether corpora will be useful to them in teaching learners and fulfilling their curricular aims, and in what ways. Hence the research question is: *How can corpora be useful in English language teaching in Norwegian schools?*

This overarching research question is supplemented by three sub-questions, corresponding to the three phases of the research project. Each phase involves exploration of a different aspect of English language teaching in Norway, to see whether teaching practice can be connected to corpus linguistics. Each of these phases corresponds to one of the three articles included in this dissertation.

The first sub-question is *How are corpora used by in-service English teachers in Norwegian schools?* As a starting point, it was necessary to discover how widespread the use of corpora is among English teachers in Norway, and of those who are using corpora, how they do so. This phase of the research obtained the perspectives of corpus-using teachers.

The second sub-question concerns the perspective of teachers who do not use corpora, but are introduced to them. The 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 corpora are not widely used among them.

The final sub-question is *How can corpora be used in the English subject in lower secondary school with the current curriculum in Norway?* The most recent English subject curriculum came into effect in August 2020. This was an opportunity to discover what corpus linguistics could offer English teachers in the context of a new curriculum and new textbooks, by creating relevant corpus exercises, adapted from textbook exercises, and influenced both by how teachers have said they use corpora (the first phase of the research) and by teacher perspectives on corpora (the second phase). Due to the wide scope of the curriculum and the wide range of textbooks, this phase of the research was narrowed to lower secondary school (years 8-10) only. One reason the corpus exercises were adaptations of textbook exercises was to build upon what teachers already work

[5] In-service 'designates 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).

with. There is a strong textbook focus in primary and lower secondary school levels in Norway: at years 5-10, 70% of surveyed English teachers said they primarily used paper-based textbooks as resources (Gilje et al. 2016: 52). A second reason the corpus exercises adapted textbook exercises was to show that corpus-based teaching materials can replace, enhance or revise textbook exercises, in light of the curriculum.

The research questions have been answered via the following methods, which are described in detail in Chapter 4. The first phase of the research was a survey of English teachers in Norway, to discover how widespread the direct applications of corpora are. From the survey, corpus-using teachers were identified, and some were interviewed about how they used corpora. The second phase involved teaching corpus use to groups of in-service teachers, and some of these teachers were then interviewed about their perspectives. In the third phase, the perspectives of both groups of interviewed teachers were taken into consideration when designing five suggested corpus exercises for use in lower secondary school.

1.5 Dissertation structure

The rest of the dissertation is structured as follows. The English subject in Norwegian schools is the context of this research, so Chapter 2 explains the English subject as it is exists in Norway, and explains how direct applications of corpora fit the curriculum. Chapter 3 provides an overview of previous international research on how to increase corpus use in schools, and what the obstacles to this are. Chapter 4 presents the methodology: the theories and principles behind the methods used. Chapter 5 summarizes the three dissertation articles. These articles and their findings are discussed in Chapter 6, which also contains a conclusion. The list of references follows, and then the three dissertation articles and their appendices. Documentation of approval of this research by the Norwegian Centre for Research Data (NSD; now Sikt – Norwegian Agency for Shared Services in Education and Research) is Appendix 1 to the overall dissertation. Lawful consent text and information letters with consent forms for research informants follow as Appendix 2.

Chapter 2 Corpora and curricula in Norway

Direct pedagogical corpus applications are warranted by the *Core Curriculum – values and principles for primary and secondary education* (Norwegian Directorate for Education and Training 2017), which provides the values and principles of education that any methods applied in teaching must conform to, and the English subject curriculum, *Curriculum in English* (Norwegian Directorate for Education and Training 2019), which describes 'the content and goals' of the subject (Norwegian Directorate for Education and Training 2017: 1). In order to provide some background about the subject first, the chapter has two main sections: 2.1 describes the English subject in Norway, and 2.2 explains how direct applications fit this subject. It was mentioned in Chapter 1 (1.1) that direct applications are explicit language learning, and both sections 2.1 and 2.2 explain how the subject is open to explicit language instruction. The research phases straddled two English subject curricula. The first two phases coincided with the subject curriculum valid from August 2013 (Norwegian Directorate for Education and Training 2013), while the third phase coincided with the subject curriculum valid from August 2020 (Norwegian Directorate for Education and Training 2019). Differences between the curricula are mentioned where relevant (both here and in Chapters 3, 4 and 6). This chapter concludes with section 2.3.

2.1 The English subject in Norway

The methods of corpus linguistics are not mentioned in the English subject curriculum for Norway's schools (Norwegian Directorate for Education and Training 2019). Yet neither are any other methods for language teaching (Fenner & Ørevik 2020: 358), or learning (Speitz 2020: 44): methods are 'to be decided at the local level' (ibid.). In other words, methods are left up to individual English teachers. The reason for this probably lies in the sources of inspiration behind the subject curriculum, most noticeably the *Common European Framework of References for languages* (CEFR) (Council of Europe 2001), and its descriptors of aims for language learning (Council of Europe 2018). The current subject curriculum (Norwegian Directorate for Education and Training 2019) and its predecessor (Norwegian Directorate for Education and Training 2006 and its revision, Norwegian Directorate for Education and Training 2013) have their roots in the CEFR. In fact, the current curriculum has been called a 'third generation document with regard to the CEFR' (Simensen 2020: 35).

The curricula include 'competence aims' that pupils are meant to achieve at various stages during their schooling. This focus on competence aims in the curricula is influenced by the CEFR (Fenner 2020: 34), and the CEFR descriptors of learner proficiency and English subject curriculum competence aims are similar (Speitz 2020: 47). In both, there is a focus on learning rather than teaching (Simensen 2020: 34), but despite this the CEFR descriptors 'do not imply any information about the processes language users go through' (Speitz 2020: 49), and the CEFR does not promote any particular methods for learning (or teaching) (Simensen 2020: 33-34).

However, both the CEFR and Norway's English subject curriculum have a communicative view of language learning (Speitz 2020: 47). In Norway, English subject curricula going back to 1987 have all been influenced by the concept of communicative competence (Simensen 2020: 32), and the CEFR contains communicative language 'competences', plural (ibid.: 33). This has its roots in the Communicative Language Teaching (CLT) approach of the 1970s, in which 'real communication always has a purpose and a function', thus teaching should focus on communicative purposes and functions (Skulstad 2020: 50). A 'weak' version of CLT has been practised in Norway, in which learning a language is not seen as being the same as learning one's first language, and 'facilitating the use of English for communicative purposes' is 'not seen as the only route to learning the language' (ibid.: 56). This means 'structural and metalinguistic aspects of language' can also be focused on in teaching, and 'teachers are often encouraged to take an eclectic approach' (ibid.). This tradition of 'weak' CLT can explain why the English subject curriculum is reluctant to prescribe specific teaching or learning methods, and methods are left up to individual English teachers.[6]

Recent theoretical developments in language learning have come from cognitive research (Newby 2020: 217-218). This has no obvious influence on the English subject curriculum, where the word 'cognitive' does not appear, yet the curriculum is compatible with these developments. Cognitive linguists emphasize the importance of attention in language acquisition, stating that 'What is attended is learned, and so attention controls the acquisition of language itself' (Ellis & Robinson 2008: 3). As a result, many cognitive linguists favour, for example, explicit grammar teaching (Newby 2020: 221-222). Attention in language is related to explicit language learning because by definition explicit learning should require more attention than implicit learning. The English subject curriculum is compatible with this. Competence aims are explicit because of their use of

[6] Additionally, competence aims in the curriculum are not directly related to the basic skills in the current curriculum (Fenner 2020: 38), but they had been in the 2013 curriculum.

'form-related terms' (ibid.: 226).[7] For example, the aim for after year 7, 'identify sentence elements in various types of sentences and use knowledge of verb conjugation and declension of nouns and adjectives…' (Norwegian Directorate for Education and Training 2019: 7) shows a focus on linguistic form. In its 'core elements' section, the English subject curriculum states: 'Language learning refers to developing language awareness and knowledge of English as a system', which includes learning 'vocabulary, word structure, [and] syntax' (ibid.: 2). The language as a system is the 'structural' aspect of language mentioned above, and evidence of the 'eclectic approach' in Norway (Skulstad 2020: 56). The eclectic approach would include explicit learning.

The quotation above from the 'core elements' reveals another influential trend in language teaching pedagogy, namely language awareness, a concept not found in the previous curriculum of 2013. Language awareness 'concerns "knowledge about" language rather than language acquisition as such' (Newby 2020: 226). The curriculum has competence aims that 'state that students should have knowledge of both their L1 [first language(s)] and English and be aware of similarities and differences between the languages' (Angelsen & Hauge 2020: 333). These aims can be found at all school levels (Norwegian Directorate for Education and Training 2019: 5, 6, 7, 9, 12). Activities for developing language awareness recommended by Angelsen and Hauge (2020: 334-336) include instruction in morphology and word learning, word order, and error identification. Once again, a focus on form can be inferred, while the communicative approach would imply a focus on meaning. This is the 'metalinguistic' aspect of language mentioned above, and more evidence of the 'eclectic approach' (Skulstad 2020: 56). The 'weak' CLT of the Norwegian tradition makes it possible to have both focus on meaning and focus on form.

There is more to the curriculum than a communicative view of language learning, an additional focus on form, and no prescribed methods. It is not the case that any method can be legitimate. The English subject curriculum does not exist in isolation, and in fact 'several approaches to teaching and learning can be inferred from the *Core Curriculum*' rather than the subject curriculum (Fenner 2020: 37). The *Core Curriculum – values and principles for primary and secondary education* provides, as its title indicates, the values and principles of education. Teaching a school subject will involve having the values and principles in mind; teaching will be approached from that angle. Any methods applied must conform to the values and principles. This means that discovering how the methods of corpus linguistics would fit into English language teaching in

[7] Newby himself advocates a 'C+C' approach (communicative plus cognitive), which focuses on meaning rather than form (Newby 2020: 218-219).

Norwegian schools involves seeing it in the light of the *Core Curriculum* as well as the subject curriculum.

The concepts 'critical thinking' and 'in-depth learning' appear in the *Core Curriculum*. Critical thinking was explained in Chapter 1 (1.3). In the *Core Curriculum,* it is said to mean 'applying reason in an inquisitive and systematic way when working with specific practical challenges, phenomena, expressions and forms of knowledge' (Norwegian Directorate for Education and Training 2017: 6), and it includes thinking critically about sources, with 'room for uncertainty and unpredictability' (ibid.). In-depth learning was new to this curriculum. The concept finds its way into the curriculum because of the research already mentioned (see 1.3). The curriculum states that the school must provide sufficient time for specialization, as well as support and guidance (Norwegian Directorate for Education and Training 2017: 7, 12, 17).

As the preceding discussion has shown, there is nothing in the English subject curriculum that prevents direct applications from being incorporated into teaching in Norway. The next section focuses explicitly on how this approach is warranted by aspects of the curricular context.

2.2 How direct applications are warranted by the English subject in Norway

The benefits of corpora in language learning were presented in Chapter 1 (1.3). This section illustrates how direct applications can also be warranted by the curricular context. Direct applications are compatible with many of the aspects of the English subject in Norway: CLT (2.2.1), explicit language learning (2.2.2), language awareness (2.2.3), concepts used in the *Core Curriculum*, namely critical thinking and in-depth learning (2.2.4), and digital skills (2.2.5), which is a 'basic language skill' in the subject curriculum.

2.2.1 CLT

Even if Norway's 'weak' CLT was disregarded as a factor, there exists an argument that direct applications' focus on linguistic form is compatible with a communicative approach. Focus on form can be part of language learning, as long as form is 'studied in context so that students can associate it with meaning and use… exactly what DDL offers' (Leńko-Szymańska & Boulton 2015: 5, who give the example of Gaskell & Cobb 2004). Corpus use is related to the concept of authentic language data, as mentioned in Chapter 1 (1.3), and authenticity is part of CLT's focus on contextual meaning (Cook 1997: 224; Gilmore 2007: 97; Buendgens-Kosten 2013: 274). CLT influences approaches to 'authentic' teaching materials (McDonough, Shaw & Masuhara 2012: 22-31), in order to provide learners more than just knowledge of language structures, and the English subject curriculum stresses 'authentic… situations', 'authentic language models' and

'authentic texts' (Norwegian Directorate for Education and Training 2019: 2, 4, 5, 7). Thus, corpora can provide the required contextual meaning.

2.2.2 Explicit language learning

The English subject curriculum's references to English as a system, and to vocabulary, word structure, and syntax, have already been mentioned (see 2.1). These references are reinforced in the competence aims. It is difficult not to see the reference to language as a system as an expectation that some language teaching will be explicit, in which case this curriculum seems open to the use of linguistic data that corpora can provide, and teachers can be open to using such material. Recent textbooks written to be compatible with the curriculum have explicit language exercises; the linguistic data of a corpus may be even more useful in introducing pupils to language elements explicitly.

2.2.3 Language awareness

The use of corpora has been promoted as something that would raise language awareness (Farr 2008: 28-30). There are studies that show that both teacher educator and pre-service teacher informants perceived teacher-corpus interaction as increasing pre-service teachers' language awareness (Breyer 2011: 149, 206; Zareva 2017: 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: 206). It has been argued that interacting with corpora would also increase the language awareness of learners (Leńko-Szymańska & Boulton 2015: 3). Language awareness now appears in competence aims at all levels in the curriculum. For example, a competence aim for Vg1 is 'use knowledge of similarities between English and other languages with which the pupil is familiar in language learning' (Norwegian Directorate for Education and Training 2019: 12).[8] Under the previous curriculum, teaching the higher years of the English subject in schools could be focused primarily on sociocultural topics (Karlsen & Monsen 2020: 138, 141, 142, 144), but the current curriculum indicates that language topics of the subject should be taught as well, which makes the use of the linguistic data that corpora can provide very relevant.

2.2.4 Critical thinking and in-depth learning

Critical thinking can apply to linguistic data in a language subject. Corpus searches present practical challenges, and there is uncertainty and unpredictability in language data. For example, it might take a pupil much thought to see a semantic pattern in the collocates of a particular

[8] A clear link between this and language awareness is made in Flognfeldt & Lund 2021: 17-18

adjective. This will not be as straightforward as patterns already given to them in a textbook (e.g. Pettersen & Røkaas 2021: 226). Pupils would be expected to reflect on corpus data, and exercise their judgement.

In-depth learning can apply to a language subject, or to cross-curricular work that includes a language subject. Introducing pupils to corpora gives them the opportunity to begin an endless investigation into words and phrases. Corpus exercises can be open-ended, if pupils are to choose their own words or phrases to investigate.

2.2.5 Digital skills

There is also a digital aspect to the English subject curriculum. The use of corpora in teaching involves digital skills, identified as one of the 'basic language skills', alongside oral skills, writing and reading (Norwegian Directorate for Education and Training 2019: 4). One of the meanings given for 'digital skills' is 'being able to use digital media and resources to strengthen language learning' (ibid.). The use of corpora, which can be considered 'digital resources' for learning, can help fulfil this.

2.3 Chapter conclusion

This chapter has shown how direct applications are warranted by both the *Core Curriculum* and English subject curriculum. This means that corpora can be used for the beneficial purposes set out in Chapter 1 (1.3). The next chapter concerns the degree to which corpora have *not* been used in schools, the obstacles involved, and suggestions for how to overcome them.

Chapter 3 Research status

Relevant previous research into direct pedagogical corpus applications in English language teaching and learning does not often touch upon primary and secondary levels of education. The previous research described below has addressed attempts to increase corpus use in language teaching and learning, or has identified the obstacles to this. The chapter has three main sections: 3.1 concerns the degree to which corpora have not been used in schools, 3.2 explains the obstacles involved, and 3.3 concerns suggestions for how to overcome them. The chapter concludes with section 3.4.

3.1 Corpus linguistics and language teaching practice

Corpora do not seem to be used much in primary and secondary English language teaching. Mukherjee (2004) wrote of the 'gap between applied corpus linguistics and the reality of English teaching', in a study that used a questionnaire to survey 248 teachers of English in secondary schools in Germany. He found that 79.4% of his in-service teachers, prior to taking a corpus workshop, agreed with the statement, 'No, I don't know anything about corpus linguistics' (ibid.: 241). He concluded that this 'illustrates the low extent to which corpus linguistics has so far had an impact on teaching practice in Germany' (ibid.: 242). In the years since, further surveys inspired by Mukherjee's study have been carried out. Heyvaert and Laffut (2008) conducted a survey of English teachers in Belgium, which also showed a low level of familiarity with corpora.[9] Callies (2019) used a questionnaire to survey German teachers of English in secondary schools, and interpreted the findings from his survey of 26 teachers as reconfirming Mukherjee's findings (ibid.: 252): only 34.6% of the teachers had 'heard of' corpus linguistics in their university studies, and only 3.8% in their practical teacher education (ibid.: 250). Mukherjee had recommended educating teachers in corpus linguistics (Mukherjee 2004: 248), and Callies noted an increased awareness among teachers of corpus linguistics in the years between his study and Mukherjee's, seeing it as due to 'a younger generation of language teachers who have been trained in the use of corpora for research purposes in their university studies', but, importantly, he observed that 'this seems not to have made a significant impact on their teaching practice' (Callies 2019: 252). Another case of low impact can be found: a survey of 100 teachers in Serbia revealed that 71%

[9] It is not clear how many teachers they surveyed.

had 'never heard of DDL' (Vitaz and Poletanović 2020: 409, 415-416).[10] In Norway, it has been acknowledged that corpora have not been used in language teaching to any great degree (Cardona, Didriksen & Gjesdal 2014: 1).

3.2 Obstacles to direct applications

In previous research, teachers themselves have suggested various reasons why direct applications are not more common. For convenience here, they are grouped into five categories: lack of usability (3.2.1), perceived unsuitability (3.2.2), lack of digital skills (3.2.3), perceived lack of teacher need (3.2.4), and teacher beliefs about implicit language learning (3.2.5). A potential obstacle not included is lack of IT infrastructure, as this problem does not exist in the Norwegian education system, where it is the norm for teachers and pupils to have access to computer equipment and an internet connection.

3.2.1 Lack of usability

The concept of usability comes from software engineering, and is 'the question of how well users can use… [the] functionality [of a system]' (Nielsen 1993: 25). The software or online interface for searching corpora can be 'difficult and troublesome' from a teacher perspective (Karlsen & Monsen 2020: 131). Popular corpus software, for example AntConc (Anthony 2022), and popular online corpus interfaces, for example English-Corpora.org (Davies 2002-), 'can serve well for research purposes, but the learning curve of the tools for students is still too steep' (Xu 2022: 21). AntConc has been evaluated positively using Nielsen's usability heuristic framework (Henry & Sheepy 2022), but it is not clear whether the intended user can be outside tertiary education.

3.2.2 Perceived unsuitability

Teachers can perceive corpora as too expensive or inaccessible for schools. This perception may be due to the paywalls of larger corpora, and/or teachers' lack awareness of other, less restricted resources (Karlsen & Monsen 2020: 134). Teachers can perceive corpora as tools for linguists only, which is understandable when many corpora were compiled as tools for linguistic research (ibid.; Braun 2007: 308).

3.2.3 Lack of digital skills

[10] Only 38% of the 100 teachers are of interest here, because they were schoolteachers (28% primary and 10% secondary). 48% were private-language-school teachers, 12% taught at universities, and 2% taught at 'specific institutions' (Vitaz and Poletanović 2020: 415).

To use computerized corpora, some digital skills are required. There are teachers who may lack these skills. In interviews with teacher educators for the secondary school system of Germany, teachers' IT skills were suggested as one of the reasons corpora are not used in language teaching; and in the same study, student teachers also had 'computer-related difficulties' (Breyer 2011: 150, 207). Digital competence can be expected to vary among Norwegian teachers too (Røkenes & Krumsvik 2016). There is a large gap between official ambitions for digital competence and reality (Norwegian Ministry of Education and Research 2022: 85). Pupils also may lack digital skills. The idea that the current generation in education would be 'digital natives' seems to apply in reality only to a minority of pupils (Bennett, Maton & Kervin 2008; Bullen, Morgan & Qayyum 2011; Schulmeister 2009): it was about 34% in a Finnish study involving 'test data describing performance-based ICT skills' (Ståhl 2017). While pupils may use social media effectively, they can have problems with many other common features of a computer, for example, Microsoft Word (Karlsen & Monsen 2020: 135).

3.2.4 Perceived lack of teacher need

Teachers may not be inclined to use linguistic materials. This may occur in the higher years of English in secondary school because teachers may become more topic-focused (e.g. on sociocultural topics) than language-focused (Karlsen & Monsen 2020: 138, 141, 142, 144). This may have been due to the previous curriculum (Norwegian Directorate for Education and Training 2013), so it could be a more easily resolvable obstacle than the first three. The current curriculum, which came into effect from August 2020 (after the present research began) indicates that language topics of the subject should be taught at every level (Norwegian Directorate for Education and Training 2019: 5-12), which makes the use of the linguistic data that corpora can provide more relevant. As mentioned in the previous chapter (2.1), the English subject curriculum refers to English as a system, and states that vocabulary, word structure, and syntax must be learnt; this undeniably adds a language focus.

3.2.5 Teacher beliefs about implicit language learning

Classroom work with corpora can be either deductive or inductive, as can be seen in literature that showcases corpus activities (Breyer 2011: 52-54; Liu and Lei 2017: 31-34; Pinto et al. 2023). Both of these ways of language learning are explicit (see 1.1), which may clash with teacher beliefs about implicit language learning (Karlsen & Monsen 2020: 139). This could also be a more easily resolvable obstacle than the first three. Teaching of the English subject in Norway is open to both 'strong' and 'weak' versions of communicative language teaching (see 1.1; 2.1), so there can still be dialogue between corpus linguists and teachers of all views. Whether this would be an

issue for other countries may depend on their curricula. Curricula elsewhere may be more compatible with a focus on linguistic data (see Braun, 2007: 310; Pérez-Paredes 2020: 75; and in an L1 context, Sealey & Thompson 2007).

The effects of these five obstacles (3.2.1-3.2.5) need to be mitigated if teachers are to use corpora. The first three obstacles could be seen as 'first-order barriers' to the adoption of technology integration, that is, 'extrinsic to teachers', and the final two obstacles could be seen as 'second-order barriers… intrinsic to teachers' (Schaeffer-Lacroix 2020: 48). However, an aspect of the third obstacle, teachers' digital skills, could be seen as intrinsic (if a teacher's level of skill is seen as intrinsic to the teacher), and an aspect of the fourth obstacle, topic-focused teaching, could be seen, at least when the previous curriculum (Norwegian Directorate for Education and Training 2013) was being used, as extrinsic (if the curriculum's requirements are seen as extrinsic), so the distinction will not be employed further here.

3.3 Suggestions for increasing direct applications in practice

There has been a body of research with suggestions for connecting corpus linguistics to language teaching practice. For convenience, they are grouped here into four broad categories of suggestions: educating teachers (3.3.1), compiling pedagogically motivated corpora (3.3.2), improving usability (3.3.3), and other suggestions (3.3.4).

3.3.1 Educating teachers about, or in, corpus linguistics.

This can be expressed as 'the need for… popularization' (Mukherjee 2004: 243), 'missionary work', 'spreading the word' to teachers (Römer 2009: 84), or 'systematically populariz[ing] corpus analysis among language teachers' (Heyvaert & Laffut 2008: 497). More concretely, educating teachers can mean anything from 'seminars, conferences and workshops' (Vitaz & Poletanović 2020: 419) to 'institutionalized teacher-training courses devoted to or featuring the applications of corpora in language instruction' (Leńko-Szymańska 2014: 261). However, a number of obstacles to this approach have become apparent, such as 'the difficulty of finding suitable corpora … lack of expertise in corpus consultation and analysis … and also problems relating to time, technical issues and expense' (Chambers 2019: 471). Workshops may be 'too modest' to be sufficient (Leńko-Szymańska 2014: 272), and even when it comes to institutionalized teacher education courses, it is not clear how much education would be enough (ibid.; Karlsen & Monsen 2020: 132).

3.3.2 Compiling pedagogically motivated corpora

This is the idea of '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: 374-5). Pedagogically motivated corpora would sidestep challenges related to using corpora originally developed for linguistic research (see 3.2.2). Yet successfully relevant corpora for Norwegian schools do not yet seem to exist: a study, designed with pedagogic purposes for secondary school classrooms in Norway, and involving a set of corpora called BACKBONE, showed that there was a need for an understanding of 'what teachers do in the classroom', and the corpora struggled to be interesting and relevant to pupils (Farr & Karlsen 2022; Karlsen in preparation).

3.3.3 Improving the usability of software

This is the idea of designing 'user-friendly concordancing software … a corpus analysis software tool for classroom use' (Breyer 2006; Breyer 2011: 207). No such software is currently available. Were it to exist, and were it to overcome the abovementioned usability challenges (3.2.1), it would be a welcome solution. However, by itself this solution would be incomplete. The corpus that is to be used with the software must still be interesting and relevant to pupils.

3.3.4 Other suggestions

The above suggestions confront us with challenges. The question arises of how corpora can be made useful in English language teaching in Norway, when teachers have not received enough education in using corpora pedagogically, relevant pedagogically motivated corpora do not yet exist, and classroom software for searching corpora does not yet exist either. Xu suggests bespoke corpora, corpus tools with better usability, and better 'integration into the overall English curriculum', through dialogue and collaboration between teachers, researchers and materials developers (Xu 2022: 20-21).[11] See Chapter 6 (6.4) for further discussion of this. Meunier (2022: 347-350) suggests that DDL can be 'revamped' through links with education, engagement with SLA theory, and broader digital competences for DDL specialists. This dissertation takes steps in addressing the first of these, links with education, by attempting to link corpora to the existing education system and curricular context. The final phase of this research aims to show how suggested corpus exercises can fit the core curriculum and English subject curriculum in Norway (see the third dissertation article).

[11] A fourth suggestion by Xu is that '[m]ore research on sociocultural and/or cognitive mechanisms should be carried out to validate the effectiveness of corpus application in English language teaching' (Xu 2022: 20), and there exists research in this expanding area (ibid.: 21).

3.4 Chapter conclusion

This chapter has described the research status in terms of corpus linguistics and language teaching practice, obstacles to direct applications, and suggestions for increasing direct applications in practice. The discussion (Chapter 6) will return in particular to suggestions for increasing direct applications in practice.

Chapter 4 Methodology

The starting point for the present study (see 1.4) was the research question, *How can corpora be useful in English language teaching in Norwegian schools?* This overarching question was broken down into the following three sub-questions: (i) How are corpora used by in-service English teachers in Norwegian schools? (ii) What do in-service English teachers in Norway find useful about corpora and what do they find challenging? and (iii) How can corpora be used in the English subject in lower secondary school with the current curriculum in Norway?

In order to answer each of these sub-questions, three corresponding phases of research were designed and undertaken. This present chapter has been structured around these phases, such that section 4.1 covers the questionnaire and follow-up interviews carried out in Phase I to answer sub-question (i), section 4.2 describes how Phase II involved a corpus course for in-service teachers and subsequent follow-up interviews in order to answer sub-question (ii), and section 4.3 covers Phase III, in which textbooks were analysed and corpus-based exercises were designed with the purpose of answering sub-question (iii). In section 4.4, the ethical considerations of the research project are described and discussed. The chapter concludes with section 4.5.

4.1 Phase I. The questionnaire and follow-up interviews

A questionnaire and follow-up interviews were chosen to answer the first sub-question, *How are corpora used by in-service English teachers in Norwegian schools?* The first dissertation article covers the following aspects of these methods: how they were undertaken, the informants, and the connection between questionnaire answers and interview questions. This section focuses on other aspects, explaining the principles behind the questionnaire (4.1.1), and the interviews (4.1.2).

4.1.1 The questionnaire, distributed March-April 2018

The questions are reproduced in Table 1 below. Five aspects of the principles behind the questionnaire are presented in this section: the length of the questionnaire (4.1.1.1), the construction of the questions (4.1.1.2), the use of branches in the questionnaire (4.1.1.3), the purpose of each question (4.1.1.4), and limitations of the questionnaire (4.1.1.5).

4.1.1.1 The length of the questionnaire

The level of awareness of corpora among teachers of English in Norway has not previously been investigated, and the reason for a widely distributed questionnaire was to attempt to gauge this level of awareness.[12] Previously, the assumption that there are not many direct applications of corpora in English language teaching in schools in Norway has relied solely on educated guesses (e.g. Cardona, Didriksen & Gjesdal 2014: 1). The starting point of this research was to gather data about it. The ambition for the questionnaire was to survey as many teachers of English in Norway as possible. Following Cohen, Manion and Morrison (2011: 264) in relation to improving response rates in a survey, it was decided that the less time it takes to fill out the questionnaire the more informants would answer it, so the questionnaire became relatively short: including all branches, there are 18 questions.

4.1.1.2 The construction of questions

The questionnaire contains questions of various types. Three questions (see Table 1) seek nominal data (potentially sorting informants into categories): the grades they teach, the counties they teach in, and their approximate ages. One question seeks ratio data (on a scale with a 'true zero'): how many years the informant has been teaching in Norway. Of the other questions, four are closed, four are completely open, and six are partially open, in that they each present a series of options plus an open option. Thus, a majority of questions are either partially or completely open. A previous study (Leńko-Szymańska 2014), which issued a questionnaire about corpus linguistics to 13 pre-service teachers, shows the advantage of asking open questions. The answers to a closed question about how familiar informants were with corpora were revealed to be 'inaccurate, vague, and or even meaningless' (ibid.: 269) when the researcher cross-referenced them with the answers to subsequent open questions. The open questions revealed the problem.

Secondly, open questions '[a]llow respondents to express themselves in their own words' (Foddy 1994: 128). The partially open questions in the present questionnaire have a series of options, and these are useful in that informants can see what kind of answers are required, or be reminded of aspects of the topic that they may have forgotten (e.g. question 5, see Table 1), but this does not mean pre-set options would represent the nuances of informants' thinking, or even include options informants would choose. Therefore, in this questionnaire, open options were included,

[12] There was an indication that awareness would not be zero. In September 2017, a 9-question pilot questionnaire was given to 15 in-service secondary-school English-teacher informants from two different further education programmes in Norway. One-third of the informants answering the question, 'Have you ever heard the term corpus? Choose from one of the answers below', chose the option 'I am fairly familiar with corpus linguistics, but I have never done any practical word with corpora.'

as there was no reason to believe that listed options were exhaustive. So, for example, a 'how' question (e.g. question 5) has the option 'In another way (please specify)'.

There are, however, pitfalls to open questions. The reason this questionnaire has only four completely open questions is to ensure respondents understand the context and what kind of answers are required (advice from Foddy 1994: 184). Another advantage of few open questions is it increases the possibility of retrieving answers from informants that correspond. Also, having too many open questions in an online questionnaire has been discouraged because it is self-completed (the interviewer cannot ask follow-up questions to confusing, ambiguous, or intriguing answers), and time-consuming for the informant (advice from Cohen, Manion & Morrison 2011: 396).

4.1.1.3 The use of branches in the questionnaire

It can be difficult to ascertain how informants understand the words used in a questionnaire. In the present questionnaire, when certain options are chosen, the informant is taken to a distinct branch of the questionnaire. For example, if 'I have already done some work with corpora' is chosen in question 9 (see Table 1), the informants are asked questions 16 and 17. This feature was influenced by two previous studies that issued questionnaires about corpus linguistics: Breyer (2009, 2011) to 18 pre-service teachers, and Farr (2008) to a mix of 25 in-service and pre-service teachers. Reading the findings of these studies, it seems open questions were the manner by which the researchers ascertained informants' meanings. For example, the question 'During the process of creating the exercise, did you find suitable corpora? Please describe your choice and give reasons' (Breyer 2009: 164) is a yes/no question with an added open instruction. If this had been a closed question, one would not discover how respondents interpreted 'suitable', a word too vague to have a standardized meaning for all informants (see Foddy 1994: 184). This influenced the present research in asking for more information, but not by increasing the number of open questions (see 4.1.1.2). Instead, informants were branched off, to be asked further questions related to the option they had chosen.

4.1.1.4 The purpose of each question

There was no space in the first dissertation article or its appendices to explain the purpose of each of the 18 questions of the questionnaire. Here, this information is presented in Table 1.

Table 1

The purpose of each question in the questionnaire

Question	Purpose	Comment
Q1. In what grade(s) do you teach English? Click on all that apply. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, Vg1, Vg2, Vg3	To see whether corpus use increases or decreases by grades taught.	Nominal data
Q2. In which county (*fylke*) do you teach? Østfold, Akershus, Oslo, Hedmark, Oppland, Buskerud, Vestfold, Telemark, Aust-Agder, Vest-Agder, Rogaland, Hordaland, Sogn og Fjordane, Møre og Romsdal, Sør-Trøndelag, Nord-Trøndelag, Nordland, Troms, Finnmark	To see whether the questionnaire reaches all of Norway, and perhaps to see whether there is geographical variation in corpus use.	Nominal data
Q3. How many years have you taught in the Norwegian school system?	To see whether there is a connection between corpus use and teaching experience.	Ratio data
Q4. What is your approximate age? Click on one of the answers below. 20-29, 30-29, 40-49, 50-59, 60+	To see whether age is a factor in corpus use.	Nominal data
Q5. How do you teach English grammar? Choose any that apply to you. I teach grammar deductively (giving rules, followed by examples), I teach grammar inductively (giving examples, from which rules can be noticed or discovered), I teach grammar communicatively (I do not teach it explicitly), In another way (please specify)	To obtain a picture of grammar teaching methods used	Partially open question

Question	Purpose	Comment
Q6. When you teach the understanding and use of English in different situations, what kind of material do you base your teaching on? Choose all that apply to you. Textbooks, Dictionaries, English usage books, Other English-language books (including literature), Newspapers/magazines, Comics, Online written material, Online quizzes, Film / TV / YouTube, English-language song lyrics, Radio / podcasts / audiobooks, English speakers in the school or neighbourhood, Field studies, Other (please specify)	To see whether the 'Other (please specify)' option might reveal corpus use.	Partially open question
Q7. Do you ever base your teaching on common mistakes that your pupils make? Yes, Sometimes, No	To determine which informants should be asked Q8.	Closed question with 3 options
Q8. When you teach based on common pupil mistakes, how do you get an overview of the mistakes? Click on anything that applies to you. Personal judgement / intuition / introspection, A collection of pupil texts, Other (please specify)	To discover whether teachers are collecting pupil texts, especially as they may not know that such collections could be called corpora. This is why one of the options was 'A collection of pupil texts.'	Partially open question, asked of the 188 informants who answered 'yes' or 'sometimes' to Q7.
Q9 Have you ever heard the linguistic term *corpus*? Choose from one of the answers below. I have never heard the linguistic term *corpus* before,	The central question of the questionnaire, with the purpose of finding out how familiar informants are with corpus linguistics.	Partially open question. Influenced by previous questionnaires for pre-service teachers, namely Breyer (2011: 162) and Leńko-Szymańska (2014: 268), in which there are gradations of familiarity with corpora.

Question	Purpose	Comment
I have heard the linguistic term *corpus* before but I have little or no idea what it is, I am fairly familiar with corpus linguistics but I have never done any practical work with corpora, I have already done some work with corpora, If you have an answer that you are sure is different to the ones above, please give it here		
Q10. Do you know what any of these terms mean? Yes, Not sure, No Collocation, Colligation, Concordance, Concordancer, Frequency list, Key word in context (KWIC), Part of speech (POS)	To see whether respondents knew corpus linguistics terminology, and to cross-reference it with the answers to Q9.	Closed question, giving 7 terms, and 3 options for each ('yes', 'not sure', 'no'). This question and Q12 & Q13 check claims of familiarity with corpora, influenced by the approach of Leńko-Szymańska (2014: 269).
Q11. Where did you encounter the term(s)?	To discover where respondents encountered terms.	Completely open question. Not asked of those who did not answer Q10, or those who did not answer 'yes' to anything in Q10.
Q12. Do you know any of these web tools? Select any you know. (SKELL, Using English, BYU corpora, BNC Simple Search)	To see whether informants can identify portals to online corpora without necessarily knowing the term *corpus/corpora*.	Closed question, showing four screenshots. See comment to Q10.
Q13. Do you know other web tools similar to the ones above? If so, please name them.	To see whether informants can identify portals to corpora without necessarily knowing the term *corpus/corpora*.	See comment to Q10. Completely open question.
Q14. Above, you answered that you have heard the linguistic term *corpus* before, but you have little or no idea what it is. Where have you encountered the term?	To discover where the term is disseminated.	Completely open question, asked of those who answered Q9 this way.
Q15. You are fairly familiar with corpus linguistics, but you have	To discover how respondents learnt about corpus linguistics.	Completely open question, asked of those who answered Q9 this way.

Question	Purpose	Comment
never done any practical work with corpora. How did you learn about corpus linguistics?		
Q16. You have already done some work with corpora. In what context did you do this? Please click on any context below that applies to you. I worked with corpora as part of teacher training, I worked with corpora in a course (not a teacher training course), I have used corpora to check acceptability of use when in doubt (or when marking), I have used corpora-based materials in my teaching, I have introduced corpora to pupils, Other (please specify)	To investigate how corpora have been used by informants. Some options imply use in teaching.	Partially open question, asked of those who answered Q9 this way. Influenced by a questionnaire about corpus linguistics issued to 26 in-service teachers (Callies 2019: 251, Table 5).
Q17. What have you used corpora for? Please click on any use below that applies to you. Grammar, Vocabulary, Spelling, Idiom, Authentic dialogue, Learner language, Other (please specify)	To investigate how corpora have been used by informants.	Partially open question, asked of those who answered Q9 saying they have already done some work with corpora.
Q18. Your experience is of interest to our researcher! Would you be willing to be interviewed about your work with corpora? Yes (+ email address), No	For follow-up interviews.	Closed question, asked of those who in Q16 responded 'corpora-based materials in my teaching', 'introduced corpora to pupils', or 'Other (please specify)'.

4.1.1.5 Limitations of the questionnaire

Two limitations of the questionnaire are discussed here. The first is a problem inherent in using questionnaires. If 'an interesting issue is mentioned, about which you would have liked to ask more, you will not know until afterwards' (Wray & Bloomer 2006: 159). Other research methods, such as interviews, are better for following up. Sometimes a pilot questionnaire can draw

unexpected information out before the final questionnaire is designed for wide distribution, but it is impossible to know what all the answers to open questions will be. In an online questionnaire, it is not possible to ask follow-up questions based on unexpected answers, and the informant will not necessarily be contactable for interview later. However, one would not want to lose the advantage of using a questionnaire: they are 'useful for surveying a lot of people in many different locations' and the 'identical format means you can easily find corresponding answers across your cohort of informants' (ibid.: 158-159), and the former advantage was crucial for the goal of trying to reach as many teachers of English in Norway as possible. A partial solution to the abovementioned problem was the next part of this phase of the project. The opportunity was given to informants who stated that they had already worked with corpora in teaching to provide an email address for a possible future interview.

The second limitation in choosing to use this questionnaire was the risk of a low response rate, which 'may skew the sample… taking the trouble to reply may reflect a personality trait that also colours the answers. Those who do not respond might have given different answers' (ibid.: 159). For example, in Callies's survey of secondary-school teachers of English in Germany, 'despite wide dissemination of a call for participation to professional organizations, teaching boards and colleagues and friends in Germany' (Callies 2019: 249, footnote 1), only 26 teachers answered the questionnaire. The intention for the present questionnaire was to reach more informants in the Norwegian setting, but participation was ultimately voluntary, and there was a risk of a low response rate. The questionnaire was launched online and publicized without knowing whether the strategy for obtaining a large number of informants would succeed. It can be said to have succeeded, but that does not mean the sample is not skewed. The number of informants, although large, was a tiny fraction of the number of teachers of English in Norway, so the findings ought to be taken as an indication of these teachers' experiences and opinions only, and are not generalizable.

4.1.2 The follow-up interviews, conducted October-November 2018

Three aspects of the principles behind the semi-structured interviews are presented here: the reason for using the method, the construction of the questions, and the limitations of the interviews.

The method of semi-structured interview, or 'interview guide approach' (Cohen, Manion & Morrison 2011: 413), was chosen because of its specific advantages. With topics decided in advance, this interview method 'increases the comprehensiveness of the data' (ibid.). Borg, in a review of semi-structured interviews of language teachers, identifies advantages: the exchange

between interviewer and informant being back-and-forth rather than 'formalized', the informants speaking from their perspective in their own words, the data having the possibility to be qualitatively rich, the direction of the conversation not being predetermined, and the informants having an active role (Borg 2015: 237).

The three interview guides are reproduced in the appendices to the first dissertation article. It can be seen that most of the questions are constructed as 'how' questions. There is one 'what' question, asked of two informants. There is one 'why' question. There is one question, 'Have you any idea why it didn't come to mind?', which, in the context of back-and-forth conversations, does not function as a yes/no question, and was not answered as one, so can be considered another 'why' question. 'How' and 'what' questions in interviews are recommended, for eliciting spontaneous descriptions (Brinkmann & Kvale 2015: 159). 'Why' questions can be seen as more problematic, as they 'may lead to an overreflected intellectualized interview' (ibid.) However, the first 'why' question, 'Why did you become an English teacher?', is not overly intellectual, and the 'Have you any idea why it didn't come to mind?' question referred back to a particular questionnaire answer informants had given, was an important inquiry into informants' reasons, and such 'questions about subjects' reasons for their actions may… be important in their own right' (ibid.), which it was.

There are two main limitations of semi-structured interviews (which, incidentally, equally apply to questionnaires). The first is that informants self-report, and they 'cannot always tell you what they actually do, only what they *believe* they do' (Wray & Bloomer 2006: 154). To address this limitation, the findings of the interviews have been presented in the dissertation and dissertation articles as what informants said they do, and not as verified observations of what they do. However, what teachers say they do is relevant, because it says something about teachers' perceptions of what they do, and probably what they think one should do, which is a way of elucidating their perceptions about the use of corpora.

The second limitation is that information received depends on the questions asked (ibid.: 155). To address this limitation, effort was made not to ask loaded or leading questions. It ought to be reiterated that the interview-guide questions were based on informants' questionnaire answers, so a question like 'How have you introduced corpora to pupils?' is not as presumptive as it looks out of context. Rather, it is based on the informant having indicated in the questionnaire that they introduced corpora to pupils. Another aspect to the limitation is that '[i]mportant and salient topics may be inadvertently omitted' (Cohen, Manion & Morrison 2011: 413), and there may be

conversation topics beyond those chosen, which the present author does not realize ought to have been included.

4.2 Phase II. The corpus education for teachers, and the teacher interviews

The process and content of the corpus education for teachers, and the interviews with volunteers from among those teachers, are described in the second dissertation article, where also information about the informants can be found. The methods were chosen to answer the second sub-question, *What do in-service English teachers in Norway find useful about corpora and what do they find challenging?* This section explains the principles behind the corpus education first (4.2.1), and then the interviews (4.2.2).

4.2.1 The corpus education for teachers, September-October 2018

In-service teachers who have not been exposed to corpora or direct applications cannot have a perspective on them, hence the introductory corpus seminars. In early September 2018, before the seminars commenced, the teachers were invited to belatedly take the questionnaire of the first phase of the research, which was still online, and seven did. To question 9, five chose the answer option 'I have heard the linguistic term *corpus* before, but I have little or no idea what it is', one chose 'I am fairly familiar with corpus linguistics, but I have never done any practical work with corpora', and one did not answer it. This seemed to support the idea that the teachers would require the seminars in order to have a perspective on the pedagogical use of corpora. Table 2 shows the contents of these seminars, and the reasoning behind each of the contents.

Table 2

The contents of the corpus seminars, and their purposes

Content of the corpus seminars	**Purpose**
Seminar one	
'Vocabulary and the Use of Language Corpora' presentation.	To link corpora to the course plan on teaching vocabulary.
Guide to LancsLex (Lancaster Vocab Analysis Tool): document.	To teach how to use an online tool for vocabulary teaching. Guides compensate for a low number of seminars, and assist memory.
LancsLex worksheet. First exercise (of two) done in the seminar.	As above.
Guide to AntConc: document and video.	To teach the use of a downloadable concordancer (AntConc) and corpus (the OANC). See the purpose of

Content of the corpus seminars	Purpose
	guides, above. Videos also compensate for a low number of seminars, and assist memory.
AntConc collocation exercise: document and video.	To teach collocation (and with two verbs useful for Norwegian learners to be able to distinguish: *teach* and *learn*). See the purpose of videos, above.
AntConc case sensitive search exercise, and an exercise with both case sensitive search and not (based on workshop exercises from Charles 2018): document and video.	To teach a feature of AntConc. See the purpose of videos, above.
AntConc wild card search exercise: document and video, but not covered in the seminar itself.	To teach a feature of AntConc. See the purpose of videos, above.
Post-seminar vocabulary assignment: adjectives exercise, using AntConc with the OANC.	To teach researching with a concordancer and corpus. The language data links to the course plan on teaching vocabulary.
Seminar two	
Assigned reading for the seminar (Hasselgård 2018).	To recall the basic concepts (corpus, concordance, frequency, teachers as corpus users, learners as corpus users).
Optional reading for the seminar (Millar & Lehtinen 2008).	To help those who want to search their own corpus (e.g. of pupil texts).
'Grammar and the Use of Corpora' presentation.	To link corpora to the course plan on teaching grammar.
AntFileConverter guide.	To teach conversion of files into plain text.
Guide to copying concordance lines from AntConc	To teach a feature of AntConc.
Two AntConc progressive aspect exercises were supplied, based on corpus exercises by Dypedahl and Hasselgård (2018: 141) and Friginal (2018: 257). The first was done in the seminar.	To link corpora to the course plan on teaching grammar.
TagAnt guide	To teach how to tag a corpus.
Guide to using a tagged corpus with AntConc	To teach how to use a tagged corpus.
AntConc passive voice exercise	To link corpora to the course plan on teaching grammar.
Guide to using a tagged corpus without tags	To teach a feature of AntConc.

4.2.2 The teacher interviews, conducted December 2018-March 2019

The method of semi-structured interview was chosen for the same reasons as in the first phase, and the same limitations are recognized (see 4.1.2). An additional advantage of this method was the ability to use corpus software and interfaces in the interview setting. Six interviews were

conducted, but it was decided not to use two of them because in both cases, the informants were substitute teachers assisting in the English classroom, so they were not English teachers with their own pupils. The interview guide for all four remaining informants is reproduced in the appendix to the second dissertation article. It can be seen that most of the questions are constructed as 'how' and 'what' questions and there are no 'why' questions, which aligns with recommendations in the literature (Brinkmann & Kvale 2015: 159). There are, however, a few yes/no questions: question 7 is 'Do you ever base your teaching on common mistakes that your pupils make?', question 23 is 'Are there computer-related challenges to using corpus methods, in your case or in the case of your pupils?', and question 31 is 'If there was a book of readymade language exercises for corpus methods, would you use it?' The reasons for the presence of these questions are given in Table 3 below. Question 7 only exists to discover whether the informant should be asked question 8, which is a 'how' question; question 23 does not really function as a yes/no question, and was not answered as one; and question 34, used in only one interview, shows the researcher thinking out loud about a later phase of the project. Another noticeably different question is question 33, which consists of a series of suggestions for why it might be useful to teach with corpora, with the question 'Do you agree or disagree?' This was used towards the end of the first interview as something to reflect upon, but it was not a necessary question, because only the in-service teachers' own perspectives answer the research sub-question.

There was no space in the second dissertation article or its appendix to explain the purpose of each of the 34 questions of the interview guides. Here, this information is presented in Table 3. Note that quotes from, and references to, the English subject curriculum mean the previous curriculum (Norwegian Directorate for Education and Training 2013), which applied at that time.

Table 3

Purposes of the semi-structured interview questions in the second phase of the research

	Question	**Purpose**
1	In what year(s) do you teach English?	Metadata.
2	For how long have you taught English in the Norwegian school system?	Metadata.
3	How many pupils do you teach English to (per class)?	Background.
4	What technology do the pupils have access to?	Background.
5	What technology do you have access to?	Background.
6	This was your first semester as a [further education] student. How was the experience?	To discover whether anything affected the informant's attitude to the corpus component of the course.

	Question	Purpose
7	Do you ever base your teaching on common mistakes that your pupils make?	To determine whether the informant should be asked Q8.
8	How do you get an overview of the mistakes?	To discover whether the informant collects pupil texts (the informant may not know that such collections can be called corpora).
9	If you collect pupil texts, how do you use them?	To discover how.
10	What do you think of the use of authentic English texts in English language teaching?	To talk about authentic texts. The English subject curriculum mentioned only authentic *situations*.
11	What do you think of textbooks in English language teaching?	To discover whether the informant relies on textbooks.
12	What are your views on digital tools? Do you use them?	To find out about the informant's competency with digital tools.
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?	To discover whether the informant links this to corpus methods.
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?	To discover whether the informant links this to corpus methods.
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?	To discover whether the informant links this to corpus methods.
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?	To discover whether the informant links this to corpus methods.
17	Competence aims after years Vg1 and Vg2. Language learning: enable pupils to 'evaluate different digital resources and other aids critically and independently, and use them in own language learning'. How do you fulfil this aim?	To discover whether the informant links this to corpus methods.
18	Having been introduced to corpus methods, how much would you say you know about them?	To discover whether the informant can meaningfully assess the usefulness of corpus methods.
19	What different users do you think corpora could have?	To discover whether the informant understands the distinction between indirect and direct applications

	Question	Purpose
		(Römer 2011: 207), though not necessarily with that terminology.
20	What do you think the advantages of corpus methods are?	To hear the informant's assessment of the usefulness of corpus methods.
21	What do you think the disadvantages of corpus methods are?	To hear the informant's assessment of the usefulness of corpus methods.
22	What do you think the challenges are to adopting corpus methods?	To hear the informant's assessment of the challenges.
23	Are there computer-related challenges to using corpus methods, in your case or in the case of your pupils?	To discover whether there are computer-related obstacles to direct applications.
24	What did you think of LancsLex?	To hear the informant's assessment.
25	What did you think of AntConc?	To hear the informant's assessment.
26	What did you think of the Open American National Corpus?	To hear the informant's assessment.
27	Did you look at the concordancer in Sketch Engine? If so, what did you think of it? [This is how Q28 was phrased in the first interview.]	To hear the informant's assessment.
28	What do you think of SKELL?	To hear the informant's assessment.
29	Did you look at Netspeak? If so, what did you think of it? [This is how Q30 was phrased in the first interview.]	To hear the informant's assessment.
30	What do you think of Netspeak?	To hear the informant's assessment.
31	What do you think of corpora as authentic language material?	To link corpora to Q10.
32	Did exposure to corpus methods make you think any differently than before? (If so, how? / If not, why not?)	To discover whether corpus methods affected the informant's thoughts about language, teaching, or anything else.
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.	To discover whether the informant agrees corpus use is important or useful.
34	If there was a book of readymade language exercises for corpus methods, would you use it?	To point to a future direction of research.

4.3 Phase III. Lower secondary textbook language exercises and the design of suggested corpus exercises

Phase III of the research was narrowed to lower secondary school (see 1.4; 5.3). It was designed to answer the third sub-question, *How can corpora be used in the English subject in lower secondary school with the current curriculum in Norway?* This section explains the analysis of lower secondary textbook language exercises (4.3.1) and the subsequent design of corpus exercises (4.3.2). The choice of online corpora for the corpus exercises, namely SKELL (Baisa & Suchomel 2014) and Netspeak, and the choice of English language topics, namely synonyms and collocations, are explained in the third dissertation article.

4.3.1 Analysis of lower secondary textbook language exercises, circa October 2021

The suggested corpus exercises were designed to correspond to textbook exercises, to make explicit that corpus exercises can be used for the same curriculum-linked language topics as textbooks. Corpus-based teaching materials can replace, enhance or revise textbook exercises, as shown in the third dissertation article. An investigation of the textbooks helped identify prominent language topics, namely synonyms and collocations, that could be targeted in the corpus exercises. The books were analysed before the topics (synonyms and collocations) were chosen for the corpus exercises. These books, published by prominent Norwegian publishers (H. Aschehoug & Co. (W. Nygaard) AS, Cappelen Damm AS, and Gyldendal Norsk Forlag AS), were lower-secondary-school textbooks linked to the curriculum that came into effect in August 2020 (Norwegian Directorate for Education and Training 2019). The ten books were from three textbook series: *Engelsk*, *Enter* and *Stages*. Each book was analysed to discover what its exercises are, what language topics are involved, and how they link to the curriculum. A guidebook for English teachers that links curricular aims for vocabulary development to vocabulary topics, and links curricular aims for grammar competence to grammar topics, was used as reference for linking textbook exercises to curricular aims (Flognfeldt & Lund 2021: 89-90, 203-204). Table 4 is an example of the analysis of one of the books: *Engelsk 8 for Cappelen Damm Student's Book* (for year 8 pupils), which indicates how each of the books was systematically analysed. Quotations in the 'Language topic' column are from the aforementioned guidebook for English teachers (ibid.), while quotations in the 'Competence aims after Year 10' column are from the English subject curriculum (Norwegian Directorate for Education and Training 2019: 8-9). Not all competence aims after year 10 have corresponding exercises in the book.

Table 4

Textbook analysis of Engelsk 8 for Cappelen Damm Student's Book *(Madsen & Mohammed-Roe 2020)*

Language topic	**Competence aims after Year 10**	**Exercises**	**Page(s) of book**
'Language typology and word order conventions Basic grammatical terminology (metalinguistic awareness)'	'explore and describe some linguistic similarities and differences between English and other languages… [the pupil] is familiar with'	Explain conjugation of adjectives; explain conjugation of regular and irregular verbs; explain comparatives; explain synonyms.	28, 95, 139, 209, 232.
'Major word classes and basic clause patterns Clause types (simple, compound, complex)'	'use knowledge of word classes and syntax'	Similar words in different word classes.	53
'Inflectional morphology Clause types and patterns Cohesive ties, paragraph structure, punctuation'	'follow rules for… word inflection, syntax and text structure'	Sentence starters	138 See also 104
'Words and lexical chunks'	'varied vocabulary'	Similar words; synonyms.	53, 216, 232-233. 187: synonyms are one of the language 'targets' of Chapter 5.
'Words and lexical chunks'	'expressions adapted to the purpose, receiver and situation'	Lexical fields.	200
'Question formation: Auxiliary *do*, question pronouns and determiners'	'ask questions and follow up input when talking about various topics…'	Write yes/no questions.	29
Verb forms, tense, concord	Related to year 7 aims	Conjugate present simple and continuous; concord with pronouns.	17, 20-22, 27-28, 43, 58, 69, 76, 95, 110-111, 129, 132, 139, 148, 156, 165, 216
Adjectives	Related to year 7 aim	Conjugate adjectives.	58-59, 86-87, 94-95, 148, 200, 208-209, 232
Articles	Related to year 4 aim	Indefinite article.	165
It and *there*	None	*It* or *there.*	149

Once the language topics for the corpus exercises (synonyms and collocations) were chosen, the relevant exercises were revisited and examined in more detail. They were categorized, into eight types of synonyms exercises and two types of collocations exercises (see the third dissertation article).

Five examples of textbook exercises, which can either be replaced, enhanced or revised with the suggested corpus exercises, are reproduced in the appendices to the third dissertation article, with the permission of the copyright holders.

4.3.2 The design of suggested corpus exercises, circa May 2022

Five suggested corpus exercises (three for synonyms and two for collocations) were designed for the third dissertation article. These can be found in the appendices to the third dissertation article. Eight types of synonyms exercises in the textbooks were categorized, and for reasons of space, there are only three corpus exercises for synonyms; the third dissertation article does, however, offer corpus-based approaches for the other five types.

Corpora as resources would be especially useful if they could enhance, or were more useful than, the resources that teachers already use. Each corpus exercise was designed on the basis of the difficulties or disadvantages of the relevant textbook exercise. The reasoning behind each exercise is presented in the third dissertation article. More broadly, the exercises were intended to align with the benefits of corpora in language learning presented in Chapter 1 (1.3). The approach also recognized the strong textbook focus in lower secondary in Norway mentioned earlier (1.4), but also theories about limitations of textbook exercises. Textbooks are considered by some commentators to be a 'constraint', providing 'a limited selection of possible ways to respond to … learning aims' (Lund 2020: 349-350). Tomlinson's guidance on materials development in language teaching states that '[m]aterials can be informative (informing the learner about the target language), instructional (guiding the learner in practising the language), experiential (providing the learner with experience of the language in use), eliciting (encouraging the learner to use the language) and exploratory (helping the learner to make discoveries about the language),' while warning that most 'commercially produced' materials, including textbooks, are only informative and instructional (Tomlinson 2012: 143). The intention for the corpus exercises, all of which involve pupil-corpus interaction, was that they would be informative, instructional, experiential and exploratory for pupils,[13] in line with benefits of corpora in language learning, and the curricular context (see Chapters 1 and 2). The corpus exercises aim to be informative in that

[13] Additional activities would be required to make the exercises 'eliciting' exercises.

they involve explicit learning (see 1.1). The corpus exercises aim to be instructional in that they reveal language in context; they provide information about conventional usage (see 1.3). The corpus exercises aim to be experiential in that they provide access to authentic language data (see 1.3). The corpus exercises aim to be exploratory when they involve inductive learning (where pupils choose which words to investigate, or where pupils are expected to discover semantic patterns for themselves); also, corpus exercises can be open-ended, which is exploratory (see 2.2.4).

In October and November 2022, the exercises were given informally to a number of pre-service teachers in two different teacher education institutions to test out. This helped with finalizing the versions that are in the appendices to the third dissertation article.

4.4 Ethical considerations

The Norwegian Centre for Research Data (NSD) approved the methods of this research project. From 1 January 2022, NSD became part of Sikt – Norwegian Agency for Shared Services in Education and Research. NSD/Sikt approvals are in Appendix 1. The 'assessment from the NSD data protection representative for research' (11 January 2018) is the initial letter of approval. This mentions only the questionnaire, but each time the Notification Form was updated online by the present author, approval was given online for the updates. This is implied in the printout of the final update, the 'assessment of processing of personal data' (2 December 2021). NSD approved the following: the questionnaire, the lawful consent on the first screen of the online questionnaire (Appendix 2), the interview guides for research phases I and II, and the information letters with consent forms for the interviews (Appendix 2). There were three different information letters with consent forms: one for follow-up interviews with questionnaire informants, one for participants in the corpus seminars (for those who wished to consent to their assignments being used in the research project, in the eventuality that assignments would be useful), and one for follow-up interviews for seminar participants. NSD also approved the operation of the questionnaire under the Qualtrics terms of service and Qualtrics privacy statement (not reproduced in the appendices).

Through the lawful consent screen of the questionnaire, and the information letters for both sets of interviews, informants were informed about their rights of participation, and the use of the data. For each interview, the informant signed the consent form.

The anonymity of questionnaire informants was ensured in the design of the questionnaire, in which no names were collected. The social media groups in which the questionnaire was

publicized have not been mentioned by name either. The anonymity of interview informants was ensured in a number of ways. The only names used have been randomly chosen pseudonyms. The names were anonymized using the random name generator at behindthename.com. The names and locations of informants' places of work have not been mentioned.

4.5 Chapter conclusion

This chapter has explained the methods of the three phases of the research project. The first phase was the questionnaire and follow-up interviews. The second phase was the corpus education for teachers and subsequent interviews. The third phase was the analysis of lower secondary textbook language exercises and design of suggested corpus exercises. The next chapter of this dissertation will present summaries of the three dissertation articles, including the findings from using these methods.

Chapter 5 Article summaries

This chapter summarizes the three dissertation articles and their findings. The first two articles are published in peer-reviewed journals. The first article (5.1) has been published in the *Nordic Journal of English Studies*, and the second (5.2) has been published in the *Nordic Journal of Language Teaching and Learning*. The third (5.3) has not been published at the time of writing. The chapter ends with an overall summary of the research project (5.4).

5.1 'Bridging the Gap from the Other Side: How Corpora Are Used by English Teachers in Norwegian Schools' (Kavanagh 2021a)

The aim of this phase of the research was to discover whether and what corpus linguistics tools and methods are being used in Norway by in-service teachers in their teaching. A countrywide survey of how familiar in-service English teachers are with corpus linguistics was undertaken, and from the survey, corpus-using teachers were identified. The rationale behind this was that collecting data from corpus-using in-service teachers might become a starting point for spreading the use of corpus linguistics among teachers, and that investigating the school context may be insightful for the corpus education of teachers, a 'bridge' from teaching practice to corpus linguistics. The research question was *How are corpora used by in-service English teachers in Norwegian schools?*

The methods used were an online questionnaire, and follow-up semi-structured interviews with corpus-using teachers (see Chapter 4). The questionnaire was answered by 210 informants. Three corpus-using teachers consented to face-to-face interviews, two teaching at upper secondary (pseudonyms: 'Joy' and 'Thomas') and one teaching at lower secondary (pseudonym: 'Lars').

Findings can be summarized under three headings: familiarity with corpus linguistics, what corpora are used for and how, obstacles to the use of corpora.

Familiarity with corpus linguistics. The questionnaire contained a question about 'the linguistic term *corpus*' (question 9, see Table 1 in Chapter 4). It was answered by 193 of the 210 informants. Informants chose from optional answers. Each option of the question showed a different level of familiarity: 15% of the 193 informants indicated they had never heard the term; 39% indicated they had heard the term but had little or no idea what it meant; 28% indicated they were 'fairly familiar' with corpus linguistics but had done no practical work with it; and 18% had done some

work with corpora.[14] (Based on answers to other questions, it can be seen that not all of the 18% were familiar with important terms in corpus linguistics.) 6.2% of the 193 indicated they used corpus-based materials in teaching or introduced corpora to pupils. These were informants interpreted as 'corpus-using teachers' for the purpose of interviews. 5.7% of the 193 (including all three interviewees) indicated they used corpus-based materials in teaching, and 3.1% of the 193 (including two of the interviewees) indicated they introduced corpora to pupils. These findings indicate that direct pedagogical corpus applications may not occur very often in Norwegian schools.

What corpora are used for and how. A majority of the abovementioned 5.7% indicated they used corpora for vocabulary, idiom, and authentic dialogue; half of them indicated they used corpora for learner language. Question 17 ('What have you used corpora for?') was vague in terms of the nature of the activity (teaching, learning, investigating, etc.). That there was no detailed follow-up question is a limitation of the questionnaire (see 4.1.1.1; 4.1.1.5). The follow-up interviews, however, collected qualitative detail about (i) the corpora that are used, (ii) the purposes corpora are used for in teacher-corpus interaction, (iii) the purposes corpora are used for in pupil-corpus interaction, and (iv) the non-use of corpora.

(i) *The corpora that are used.* Joy indicated she uses online corpora SKELL (Baisa & Suchomel 2014), Netspeak, and GloWbE (Davies 2013). Thomas indicated he used 'COCA or something like it', most likely something from English-Corpora.org (Davies 2022-), but stopped using it. Lars uses COCA (Davies 2008-).

(ii) *The purposes corpora are used for in teacher-corpus interaction.* Joy uses SKELL and Netspeak for vocabulary, specifically to check collocations for giving feedback. She uses GloWbE for checking spelling, and for preparing teaching on varieties of English. Lars uses COCA for vocabulary, specifically to prepare teaching on collocations and to check idioms; and for checking acceptability of usage more generally. Thomas was not using corpora anymore, but had used English-Corpora.org for vocabulary, specifically to obtain frequency information to show pupils.

(iii) *The purposes corpora are used for in pupil-corpus interaction.* Joy has shown pupils how to use Netspeak for vocabulary (collocations and idioms) and grammar (verb forms), and SKELL for vocabulary (idioms). Lars has shown pupils how to use COCA for vocabulary (checking acceptability) and grammar (suffixes).

[14] 18% is 34 informants. The present author is often asked about a piece of data not included in the article, namely the age profile of the 34. It is as follows: 8 of these informants were 20-29 years of age, 15 were 30-39, 7 were 40-49, 4 were 50-59, and none were 60+.

(iv) *The non-use of corpora.* None of the three interviewees considered investigating concordances of learner texts.

Obstacles to the use of corpora. The first obstacle identified in the interviews was 'differences between school levels'.[15] In upper secondary, there was less focus on the English language in the English subject, and more focus on communicating about sociocultural issues.[16] In lower secondary, there was more focus on language, but concepts which corpora are useful for might be less relevant to pupils at this level,[17] and Lars only showed COCA to his best-performing pupils. The second obstacle identified was lack of usability: Thomas and Lars criticized online corpus interfaces as confusing, messy and slow. The third obstacle identified was an aspect of the perceived lack of teacher need (see 3.2.4), but one not found in previous research: teachers may not be inclined to consult corpora if they perceive that they already know the pattern of their pupils' language mistakes, due to the repetitive nature of their work with pupils at the same level over a number of years.

In concluding remarks, the article suggests that a starting point for corpus use among teachers may be to teach the tools and methods that seem to be already working for some in-service teachers.

5.2 'Norwegian in-service teachers' perspectives on language corpora in teaching English' (Kavanagh 2012b)

From the findings of the previous article, it seemed that direct applications do not occur often in the English subject in Norwegian schools. It is important therefore that corpus linguists understand pedagogical needs. To this end, the second article explores the in-service teacher perspective. The research question is *What do in-service English teachers in Norway find useful about corpora and what do they find challenging?*

[15] This is as it is referred to in the article, but this obstacle is presented as 'perceived lack of teacher need' in this dissertation (3.2.4).

[16] The current English subject curriculum may have caused this to change; the old curriculum was in effect at the time of the interviews (see 3.2.4).

[17] The article gives collocation as an example of such a concept (Kavanagh 2021a: 17), and this was true for the old English subject curriculum that applied at the time of the interviews. The current curriculum adds a competence aim after year 10 regarding 'a varied vocabulary and idiomatic expressions adapted to the purpose, receiver and situation' (Norwegian Directorate for Education and Training 2019: 8). This aim is related to words and lexical chunks (Flognfeldt & Lund 2021: 89), and collocations are part of this element of vocabulary teaching, e.g. collocations exercises are found in year 10 textbooks for the current curriculum (Diskin & Winsvold 2021: 155; Pettersen & Røkaas 2021: 226; see the third dissertation article).

The methods used were corpus seminars integrated into a language course that is part of a further education programme for in-service teachers, and follow-up semi-structured interviews with some of the teachers (see Chapter 4). The corpus seminars were attended by 45 in-service teachers of English from two separate semester-long language courses that focus on grammar, pronunciation and vocabulary. There were two groups: 25 participants were primary and lower secondary school teachers (years 5-10), and 20 participants were secondary school teachers (years 8-13). The article reports on interviews with four teachers. The first interview as conducted through Skype and the informant (pseudonym: 'Ebba') taught at primary level. The other three interviews were conducted face-to-face, with one informant (pseudonym: 'Amanda') teaching primary, one (pseudonym: 'Rebekka') teaching lower secondary, and one (pseudonym: 'Katerina') teaching vocational upper secondary. Findings can be summarized under two headings: usefulness of corpora in teaching, and challenges of using corpora in teaching.

Usefulness of corpora in teaching. Ebba, Rebekka and Amanda thought Netspeak would be good for pupil-corpus interaction. Ebba also thought SKELL would be good for pupil-corpus interaction. The main finding is that informants found corpora useful for teaching and learning vocabulary. Amanda and Katerina saw the benefits of corpora for frequency information. Amanda praised the vocabulary tool LancsLex for that. In the previous article, in-service teacher informants had found corpora useful for more than just vocabulary; they found them useful for grammar, checking acceptability of usage, and varieties of English. The difference between the two groups of informants may be that the second group had less experience with corpora.

Challenges of using corpora in teaching. Challenges raised by the informants were: usability, computer and IT challenges, learner-corpus interaction challenges, and lack of teacher need.

In relation to usability, Rebekka thought offline concordancer AntConc (Anthony 2022) was cumbersome, confusing, and not intuitive. Katerina described it as complex. Both Rebekka and Amanda thought AntConc was more time-consuming than online corpus interfaces. Katerina also criticized the usability of the online interfaces, comparing them unfavourably with web browsers.

In relation to computer and IT challenges, Ebba and Rebekka thought that a corpus, being an electronic database, would be challenging or unenjoyable for teacher colleagues. They perceived a lack of IT skills among teachers. During the seminars, the present author also observed teachers having problems during the seminars, with downloading files, unzipping files, and laptop power cables.

In relation to learner-corpus interaction challenges, Ebba, Rebekka and Katerina thought AntConc unsuitable for schools, referring to the complexity of both the software and the concordance lines for pupils. Rebekka was more optimistic about SKELL, but only for 'the higher levels' because she thought it could not be used in year 8, the reason being that at that level pupils still had trouble using a dictionary as instructed. Katerina thought SKELL unsuitable at all levels because she found its example sentences to be context-free and not understandable. Another kind of learner-corpus interaction challenge was lack of pupil interest in language. While Rebekka thought year 8 pupils could use Netspeak, she did not believe they would all be interested. This echoes an opinion of the informant Joy in the first article, that the average pupil is not interested in language (Kavanagh 2021a: 15).

In relation to lack of teacher need, Amanda thought there was no need for English language corpora at primary level because pupils only write short texts at that stage. Rebekka could not see the use for corpora before year 10, perceiving pupil language mistakes as 'obvious' to teachers in the lower years.

The same challenges were found in previous research. The article mentions a number of solutions that have been suggested for addressing the challenges: corpora to meet pedagogical needs and curricular requirements; the design of a concordancer for the classroom; and the improvement of teachers' and pupils' digital skills (see Chapter 3). The article suggests that it would be fruitful to discover how corpora, even if that involves only the basic use of Netspeak and SKELL, can help in the move towards explicit language learning found in the new[18] English subject curriculum.

5.3 'Corpus Exercises for Lower Secondary English' (Kavanagh in preparation)

The aim of this phase of the research is to show how suggested corpus exercises can fit the current curriculum for lower secondary school (years 8-10) in Norway. The article suggests corpus exercises that teachers of English in Norway may find accessible, and that relate to the curricular context. The research question is *How can corpora be used in the English subject in lower secondary school with the current curriculum in Norway?*

The reason the research focuses on lower secondary school is that there was insufficient time and space for a focus wide enough to embrace all of the curriculum, and textbooks for every school level. Also, lower-secondary curricular aims, and the lower-secondary teachers interviewed in the

[18] That is, the current curriculum. The old curriculum was in effect when the research presented in the second dissertation article was undertaken.

previous two articles, seemed to have comparatively more to say about language work in the English subject.

The corpus exercises were designed for learner-corpus interaction. They involve hands-on computer use by pupils. The exercises have both deductive and inductive aspects. Guidance from the teacher need not be absent in hands-on interaction (see below). To use the exercises with pupils, the teacher does not need to be a corpus linguist. The exercises are presented step-by-step, with computer screenshots. The exercises adapt exercises from lower secondary school textbooks; that the exercises correspond to textbook exercises shows that corpora can link to the curriculum in a similar way to the textbooks, or even more so.

The design of the exercises takes into consideration five obstacles to direct applications discussed in the previous two articles and in previous literature: (1) usability, (2) perceived unsuitability, (3) lack of digital skills, (4) perceived lack of teacher need, and (5) teacher beliefs about implicit language learning. The choice of corpora (SKELL and Netspeak) is intended to mitigate the effects of the first three obstacles, and the current curriculum may mitigate the effects of the last two (see 3.2.4; 3.2.5). Direct applications are warranted by both the core curriculum and subject curriculum (see Chapter 2).

The language topics in the exercises are synonyms and collocations. These were chosen because both are vocabulary topics (previous studies, including the previous two articles, show teacher awareness that corpora can be used for vocabulary teaching and learning), both topics relate to the curriculum, and exercises for both appear in lower secondary school textbooks.

The article analyses the synonyms and collocations exercises found in nine lower secondary textbooks. The suggested corpus exercises can replace, enhance or revise the textbook exercises. Three synonyms corpus exercises relating to three types of synonyms textbook exercises are provided. Two types of collocations exercises are found in the books, and one corpus exercise relating to each type is provided.

Synonyms exercises. Exercise 1 (Appendix 2A to the article) is a corpus exercise for years 8-10. Pupils are asked to find synonyms for words. The 'Similar words' and 'Examples' features of SKELL, and the 'finds similar words' feature of Netspeak, are used. Exercise 2 (Appendix 2B to the article) is a corpus exercise for year 8. Pupils are asked to collect examples of synonyms. The same features of SKELL and Netspeak as Exercise 1 are used. Exercise 3 (Appendix 2C to the article) is a corpus exercise for year 10. Pupils are asked to choose words that can replace the verb *say* and to use the verbs in sentences. The 'Examples' feature of SKELL is used. Using corpus

exercises instead of textbook exercises for synonyms solves difficulties such as where to find synonyms, which synonym to use, and how to use it.

Collocations exercises. Exercise 4 (Appendix 2D to the article) is a corpus exercise for year 10. Pupils are asked to find out what 'expressions' (e.g. *throw a switch* and *surge of power*) mean, from the contexts in which they are used. The 'Examples' feature of SKELL is used. Exercise 5 (Appendix 2E to the article) is a corpus exercise for year 10. Pupils are asked to find out what kind of words go together with certain adjectives. The 'finds one word' feature of Netspeak is used. Using corpus exercises instead of textbook exercises for collocations solves difficulties such as lack of examples or evidence, and removes potential disadvantages such as semantic patterns being provided to pupils in advance.

The teacher has the responsibility of introducing SKELL and Netspeak to pupils, and the article suggests some activities. The teacher may also guide pupils through exercises. Where pupils are expected to discover semantic patterns for themselves, the teacher could lead discussion, once pupils have gathered enough evidence. The teacher does not have to (i) know more than the most accessible features of SKELL and Netspeak, (ii) assess the suitability of materials more than would be necessary with a textbook exercise, or (iii) design corpus exercises.

The article concludes that researchers could devote time to designing hands-on, step-by-step exercises, or other corpus-based teaching materials, for specific curricula, which can then be made available to teachers.

5.4 Overall summary of the research project

The questionnaire from the first phase of the research revealed that only 34 out of 193 teacher informants had done some work with corpora, indicating scant evidence of direct applications in Norwegian schools. The subsequent interviews with three corpus-using teachers showed that they used online interfaces, such as SKELL, Netspeak, COCA and GloWbE, rather than offline concordancing software. The interviewees' teacher-corpus interaction was for reference, and for creating vocabulary and varieties-of-English exercises. Pupil-corpus interaction was encouraged by two of them. In the second phase of the research, four in-service teachers who were introduced to corpora and interviewed thought corpora were useful for teaching and learning vocabulary. What they found challenging confirmed previous research. Additionally, the interviews revealed a perceived lack of need for corpora before year 9 (Kavanagh 2021b: 100). The third phase of the research suggested how corpora can be used in the English subject in lower secondary school (years 8-10) with the current curriculum in Norway, by creating relevant

corpus exercises, adapted from current textbook exercises, and influenced by how teacher informants said they use corpora (Phase I) and teacher informant perspectives on corpora (Phase II). This summary of the research project serves as the starting point for the discussion in the next chapter.

Chapter 6 Discussion

This chapter is a discussion of the implications of the findings of this research project, as set out in the three articles, and summarized in Chapter 5. First, discussion focuses on how teachers, even corpus-using teachers, do not seem to mention corpora as a resource (6.1). The next section (6.2) focuses on implications linked to the content of the English subject, and the following section (6.3) discusses including primary school pupils (years 1-7) in direct pedagogical corpus applications. 6.4 discusses ongoing suggestions for increasing direct applications in practice, while 6.5 adds this dissertation's own suggestion, namely a book or collection (a 'toolbox') of corpus exercises for the English subject. The conclusion (6.6) discusses this research project's contribution to the field, its limitations, and future directions for research.

6.1 Corpora as an unmentioned resource

The focus of this dissertation is the use of corpora from the perspective of teaching practice. The findings of the first phase of this research project indicated that corpora, in relation to teaching, do not seem to have been on the minds of teacher informants, even when teachers were familiar with corpora. This was apparent in the answers to question 6 in the questionnaire: 'When you teach the understanding and use of English in different situations, what kind of material do you base your teaching on? Choose all that apply to you.' In response, none of the informants mentioned corpora (see the first dissertation article, Kavanagh 2021a: 10). The three corpus-using interviewees, who evidently considered corpora as an aid to their teaching, later gave reasons why they did not mention corpora in answer to this question (first dissertation article, Kavanagh 2021a: 12). One said he shows corpora to individual pupils, which did not for him count as material you 'base teaching on'. The question had made him think only of material for the whole class, 'all of them together', even though showing material to individual pupils is still teaching. The second informant said, 'I just think of it [corpus use] more in terms of dictionaries really.' Looking at her answer to questionnaire question 6, she chose 'dictionaries' as material she 'based teaching on'. For her, corpora were included in this. The third informant engaged in teacher-corpus interaction, but he did not think of corpora as something he 'based teaching on', because he was not engaged in showing corpora to pupils. In each case, there was corpus use, but 'material you base your teaching on' meant different things to different informants. This multifariousness is not an obstacle to corpus linguistics. Rather, it is evidence of direct applications happening, whether corpora are used for individual pupils, called 'dictionaries',

or used by teachers alone. The usefulness of corpora is evident, even if corpora do not spring to teachers' minds when answering a questionnaire.

The informants' answers to question 6 were also a snapshot of their opinions at the time of a previous subject curriculum. In the current curriculum, the usefulness of corpora as a resource may be more perceivable. As argued above (2.2), direct applications are warranted by the English subject curriculum and the core curriculum, because such applications relate to CLT (e.g. authentic language data), explicit language learning, language awareness, critical thinking (e.g. drawing conclusions from data), in-depth learning (e.g. open-ended investigation) and digital skills (e.g. pupils expected to use different digital resources). 'A subject curriculum is an all-important document that teachers need to know' (Speitz 2020: 40), and this applies equally to the core curriculum. When teachers seek materials, tools and methods for fulfilling these aspects of the curricula, language corpora can be attractive to them. The next section will comment on what areas of the English subject corpora can be linked to.

6.2 The content of the English subject

The system of the English language (its grammar, phonology, etc.) is part of the content of the English subject. There is more mention of this system in the current subject curriculum than in the previous one. For example, in its 'core elements' section, the current subject curriculum states: 'Language learning refers to developing language awareness and knowledge of English as a system', which includes learning 'vocabulary, word structure, [and] syntax' (Norwegian Directorate for Education and Training 2019: 2). This makes explicit teaching about language structures more likely (see Chapter 2). Under the previous curriculum, by upper secondary school it seemed this aspect of the subject had been downplayed. As an upper-secondary teacher informant put it, '[t]here are so many other things we have to teach we do not really get to go into language' (first dissertation article, Kavanagh 2021a: 17). Upper-secondary teachers focused on pupils being able to communicate about culture, society and news items (see 2.2.3; 3.2.4). In the previous subject curriculum, 'culture, society and literature' was a main 'subject area' with competence aims at every level (Norwegian Directorate for Education and Training 2013). The current subject curriculum has such aims, but not as a distinct 'subject area' (Norwegian Directorate for Education and Training 2019). All of the aforementioned changes mean that even in upper secondary, when pupils' communicative competence is high, pupils can still 'get to go into language', and explicit language teaching (and language corpora) can play a role at this level.

Compellingly, even if the system of the English language was still downplayed as an aspect of the subject, corpus linguistics could still have a role. Corpora can also be relevant for the sociocultural topics that pupils engage with in upper (and lower) secondary. For example, Palma (2022) describes how corpora can be used for teaching culture and intercultural communication, highlighting two case studies. The first was of refugee representation in a corpus of British newspapers, and the second was of gender minority representation in a corpus of British advice threads. Both case studies illustrated that cultural nuances and societal biases can be found in corpora of authentic examples. This kind of work is advantageous to pupils in all the same ways as work with aspects of the system of the language, because it involves authentic language data, language awareness, critical thinking, in-depth learning, and digital skills.

6.3 Primary school pupils and direct applications

In terms of teacher-corpus interaction, some teacher informants did not perceive the usefulness of it for the early years of education, because of pupil mistakes that they perceive as 'obvious', and because pupils do not write long English texts in primary school (first dissertation article, Kavanagh 2021a: 18; second dissertation article, Kavanagh 2021b: 100). Yet there are many language topics in the competence aims after years 2, 4, and 7 (Norwegian Directorate for Education and Training 2019: 5-8; Flognfeldt & Lund 2021: 89-90, 203-204). Corpus data can help teachers with their language awareness (see Chapter 2), and corpora can also be used to check acceptability of usage (see 1.3). There is even a competence aim for pupils after year 2 to 'find high frequency words and phrases in different types of texts' (Norwegian Directorate for Education and Training 2019: 5), which seems to assume year-1 teachers have knowledge of the concept of frequency information (see 1.3).

Pupil-corpus interaction is also possible in primary school. In the second phase of this research, an informant related that when her year-8 pupils search for a word in an online dictionary like Ordnett, they 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 (Baisa & Suchomel 2014) examples might therefore be difficult (second dissertation article, Kavanagh 2021b: 100). This raises the question of whether pupil-corpus interaction might be more appropriate for year 9 onwards, but the answer is that this is not necessarily the case. The interview took place when the previous subject curriculum was in effect. In that curriculum, there was no mention of dictionaries (Norwegian Directorate for Education and Training 2013), while in the current subject curriculum, dictionary competence is expected by the end of year 4 – 'explore different dictionaries and how they can be used in language learning' – and by the end of year 7 it is

expected to be more advanced: 'use digital resources and different dictionaries in language learning, text creation and interaction' (Norwegian Directorate for Education and Training 2019: 6, 7). Dictionary familiarity is now supposed to be developed at an earlier stage than year 8. Provided corpora do not have insurmountable usability challenges for pupils, they too can be 'digital resources' at primary level. There are not many studies on using corpora with primary schoolchildren (Crosthwaite 2022: 378), but a multimodal corpus tool has been developed for primary school English learners in Japan (Hirata 2020). There is no reason to believe this will not be an expanding area.

6.4 Ongoing suggestions for direct applications in practice

Suggested solutions from previous research for increasing direct applications in practice were presented earlier (3.3). This section discusses how the present research is related to this ongoing conversation. Suggested solutions are dealt with below one by one. The main three are: more education (6.4.1), new corpora (6.4.2), and new software (6.4.3). Suggested collaboration between teachers, researchers and materials developers is also discussed (6.4.4).

6.4.1 Type of solution: more education

Earlier, the idea of educating teachers about or in corpus linguistics was presented, and it was unclear how much education would be enough to make this work (3.3.1). This is an ongoing area of research (see Allan 2023). However, a recent recommendation for practice is to use more accessible tools not requiring previous training (Meunier 2022: 352-354). The exercises designed as part of the present research (see the appendices to the third dissertation article; see also 6.5) can be seen as an attempt to do the latter. It is necessary to consider teachers who have no previous training with corpora, because the questionnaire findings in the present research indicate that there are more teachers of English in Norway who are unfamiliar with corpora than teachers who are familiar with them (see the first dissertation article, Kavanagh 2021a: 11, 29).

For a teacher to also become a corpus linguist, it must be remembered that, as a previous study concluded, 'only extensive exposure to corpora by future teachers coupled with suitable teacher training in the applications of corpora in language education may bring a substantial change in the scope of corpus use in language classrooms in the wide educational context' (Leńko-Szymańska 2014: 260). Were teachers corpus linguists, and were they also to have a high level of digital competence (see 3.2.3), this would undoubtedly benefit pupils, but at the present moment it does not seem likely that pupils have, or will have, such teachers en masse in Norway. This means the

'more education' solution will be very slow to take effect, while the idea of using more accessible tools not requiring previous training can be acted upon much more promptly.

6.4.2 Type of solution: new corpora

Earlier, the idea of designing new, pedagogically motivated corpora was presented, yet no corpus of this type has so far has successfully related to what teachers undertake in the classroom in Norway (see 3.3.2). The idea of new corpora remains a point of discussion. A recent historical overview of using corpora in English language teaching ends with some 'insights for future work', and the very first of these insights is that more corpora are needed that 'match learners' ages, current language proficiencies, and even their individualized learning needs' (Xu 2022: 20). Educators may be waiting some time for these corpora. It may be that funding will never materialize for bespoke corpora to be developed for the English subject curriculum in Norway. Pedagogical activity related to corpora will therefore be linked to corpora that already exist. For example, in the suggested exercises in the third phase, these were online corpus interfaces SKELL and Netspeak. Pedagogical activity could also involve 'Do-It-Yourself' (DIY) learner corpora (Millar & Lehtinen 2008), or 'Micro Corpora' (Palma 2022), which can also be DIY. The idea behind a DIY learner corpus is that a teacher can put together 'structured collections of language produced by language learners', that is, by their pupils (Millar & Lehtinen 2008: 61). This language data can 'encourage teachers to reflect on their own teaching, investigate and, where possible, address the causes of phenomena observed in the corpus', or it could be used in the classroom for pupils to notice the gap between the input and their own language output (ibid.: 67). The definition of a Micro Corpus is 'a small collection of representative texts gathered to conduct a concise analysis of a specific linguistic issue in context' (Palma 2022: 121). The corpus would be collected by a teacher or pupils for 'precise research questions', and Palma's idea is that such corpora would be used to 'develop critical thinking on ICC [intercultural competence] and societal issues' (ibid.).

Rather than waiting for bespoke corpora to surface, existing corpora, DIY corpora, and Micro Corpora can become associated with the classroom and curricula of today. The suggested exercises are intended to show how, for example, SKELL and Netspeak can be used in the current educational context in Norway, and these are by no means the only suitable corpora.

6.4.3 Type of solution: new software

Another idea presented earlier was designing concordancing software specifically for classroom use (see 3.3.3). This would solve the obstacle of poor usability of software for teachers and

pupils. However, it is a tool for classroom use that does not yet exist. The abovementioned historical overview of using corpora in English language teaching has, as another of its 'insights for future work', the insight that 'More user-friendly corpus analysis tools need to be developed' (Xu 2020: 20), which shows that this is an ongoing concern, with no solution in sight. This is something that may never be developed for the English subject curriculum in Norway. So rather than await such a development, pedagogical activity related to corpora must rely on software or online interfaces that exist now. Ways must be found to use these for the contemporary classroom and curriculum. The exercises in the third phase used example interfaces that were thought to have fewer usability problems than other interfaces (see the third dissertation article).

6.4.4 Type of solution: collaboration

Concern has been raised concerning how corpus use fits actual curricula, as illustrated by the following two quotes:

> Very few DDL studies provide information on how the activities fit in the broader curricular context in which they are carried out and this often leads to weak instructional designs… The lack of explicit verbalisation on how DDL activities meet curricular demands or expectations is one of the reasons for its lack of uptake in teaching contexts other than university-level courses (Meunier 2022: 348).

> …the lack of integration into the overall English curriculum might be a major drawback of the corpus approach to language teaching (Xu 2022: 21).

These scholars' concerns chime with the present research; the exercises were designed to fit the current curriculum in Norway. It is, however, possible that design of these exercises is only a partial solution, and that a greater connection with teachers and materials developers might produce more targeted and appropriate corpus exercises. One of the scholars quoted above recommended: 'More dialog and collaboration with language educators, practitioners, and ELT materials developers should be encouraged in order to bridge rich language data, and diversified, as well as individualistic learning needs' (Xu 2022: 21). Corpus exercises could be designed within a collaborative framework. It would not then require one person to have multiple skill sets in order to produce them. Instead, they could emerge from the combined efforts of different professionals: the linguist/researcher, the teacher and the materials developer. This may solve the problem of teachers not being corpus linguists, or linguists not understanding the classroom, or any gap in understanding that materials developers may have.

All the same, it is not conclusive that collaboration is the formula for increasing direct applications. A problem that can occur in collaboration is one of 'ownership'. For example, in a researcher-teacher collaboration using the BACKBONE corpora, 'the teacher … felt little ownership over the tasksets, since much of their design had been done by the researcher' (Farr & Karlsen 2022: 335). Because of this problem, the present author's suggested solution for the Norwegian context does not specifically require collaboration (although it may benefit from it). The suggested solution is presented next.

6.5 A collection of corpus exercises for the English subject

One of Xu's 'insights for future work' is '[e]xperimentation into the integration of corpus resources with overall teaching objectives' (Xu 2022: 20). The exercises of this research were designed to be an example of integrating corpus resources into teaching objectives. In the absence of being able to provide more education, new corpora, or new software, designing corpus exercises might be a fertile approach for spreading direct applications. A requirement would be that the corpus exercises are made attractive to teachers within the Norwegian school context. Here it is suggested that teachers could be reached through the type of material they already tend to consult. Language corpora can fulfil curricular aims, but not all teachers are directly engaged in searching for materials, tools and methods for fulfilling these aims. Many teachers rely on textbooks to interpret the curriculum for them (Rødnes & de Lange 2012). Because of this, a close link between textbooks and corpus exercises could be a fruitful endeavour. If existing materials that teachers already rely on can directly point them onwards to more authentic language data, opportunities to investigate data and draw conclusions, and opportunities to exercise digital skills, then corpora could seem accessible and facilitative of teachers' needs. The present corpus exercises were based on existing textbook exercises, and those textbooks were in turn published to be used with the current curriculum. There is a strong textbook focus in primary and lower secondary school in Norway (see 1.4), and the present survey data showed that 180 out of 198 informants based teaching on English usage on textbooks.[19] Given this focus, a textbook would make a suitable avenue through which direct applications could take place in the English subject in Norway.

A book of exercises, or a digital collection of the same, could be designed taking the classroom as the starting point. Corpus exercises can be best presented to teachers if designed to fulfil their curricular and classroom objectives. The benefits of corpora in language learning, namely

[19] 198 informants, out of the 210 informants who answered the questionnaire, answered question 6 (see the appendices to the first dissertation article, Kavanagh 2021a: 27).

frequency information, information about conventional usage, practice in critical thinking, access to authentic language data, and in-depth learning (see 1.3), indicate what kinds of exercises the collection could contain. Providing teachers with this material means that teachers themselves do not have to be corpus linguists or materials developers.

A free e-book containing lessons for classroom corpus use has been published which perhaps provides a partial model for what is suggested here (Pinto et al. 2023). Its lessons are for various languages, it uses various corpora, it involves both inductive and deductive language learning, and there are both hands-on (internet) and hands-off (no internet) activities. A more applicable version of this e-book would be one consisting of bespoke exercises matched to a specific curriculum and/or pupil level. To build interest among English teachers in the Norwegian context, each lesson in this kind of book would need to specify which school grade(s) it is suitable for, and what the curricular aims or principles are. These last aspects of the model ought to be stressed: it is difficult to work out from the e-book (Pinto et al. 2023) which pupils in the Norwegian system the English exercises would be suitable for, and how the lesson goals would fit the curriculum.

The type of book or collection contemplated here may provide a solution to the 'ownership' problem affecting collaboration, mentioned in the previous section. Even though a book of corpus exercises are exercises designed by someone other than the teacher, it is something the teacher can work with without its author; it has no 'person' attached to it. Relating to an object is different than relating to a person, in that a relation between people, without an object they can relate to instead, is at risk of becoming an instrumental subject-object relation (Skjervheim 1957, 1996). The teacher can pick and choose from the exercises, using it freely, like a toolbox.

A 'toolbox' of corpus exercises is a solution to the obstacles presented in Chapter 3, namely lack of usability, perceived unsuitability, and lack of digital skills (3.2.1-3.2.3). Such a collection would aim to avoid very challenging software or interfaces, to use corpora that are free and accessible, to not give the impression that the approach is only for linguists (the term 'corpus exercises' does not even have to be used), and to avoid requiring prior teacher training to use it. The latter aim may be the most novel thing about this approach: the idea is to present teachers with solutions for their pedagogical needs, solutions which happen to involve corpus linguistics. The teachers would not have to be trained in corpus linguistics themselves, merely choose from the available exercises.

Two other obstacles to direct applications were mentioned earlier, namely perceived lack of teacher need, and teacher beliefs about implicit language learning. In both cases, the current

subject curriculum may make these obstacles resolvable (see 3.2.4; 3.2.5). The effect of the first of these obstacles might also be mitigated if the proposed book contains corpus exercises for sociocultural topics (see 6.2).

6.6 Conclusion

This final section concludes the discussion, and the dissertation. The discussion is summarized in 6.6.1. The research project's contributions to the field are discussed in 6.6.2, and its limitations are discussed in 6.6.3. In 6.6.4, future directions for research are proposed. Finally, 6.6.5 returns to a notion introduced earlier when introducing the research question (1.4), that of a 'bridge' from corpus linguistics and teaching practice, although this time it is the other way around: from teaching practice to corpus linguistics.

6.6.1 Summary

The research question was *How can corpora be useful in English language teaching in Norwegian schools?* This research project has investigated teacher perspectives of direct applications through a questionnaire, corpus seminars, and interviews. Corpus exercises were designed, as part of a suggested way forward. As this dissertation has argued, there are benefits of direct applications (Chapter 1) and they fulfil curricular requirements (Chapter 2); there are obstacles to their use (3.2) and suggestions for increasing their use (3.3). This discussion has suggested its own solution, a more specific suggestion for the immediate Norwegian curricular and classroom context. Using the classroom as the starting point, a book or collection of corpus exercises, based on the existing level of teacher corpus literacy, existing corpora, and existing software, could be designed and promoted to teachers as a way of fulfilling their curricular and classroom aims. The e-book by Pinto et al. (2023) is a partial example of this way forward, and their work will probably not be the last of its kind. The suggestion here is for something that is far more bespoke than that, tied to a specific curriculum and specific learners as much as possible.

The first phase of the research discovered, and interviewed, corpus-using teachers, and to a certain degree answered the question of how corpora can be useful in English language teaching in Norwegian schools, but the suggested corpus exercise collection is a proposal for how corpora can be useful to more teachers than those few corpus-users, and is a more complete answer to the question.

6.6.2 Contributions to the field

Taking the research question as being answered, this section considers what this research project has contributed to the field. The first contribution consists of the questionnaire findings, which

indicate that there are not many direct applications in Norway. There had been no empirical data which established this before. The questionnaire findings are thus useful for researchers approaching this topic in the Norwegian educational context. They may also be useful elsewhere, in that the sample size of informants is greater (relative to population size) than that of relevant past surveys of teachers (Mukherjee 2004; Callies 2019). In order to spread the benefits of direct applications, some knowledge of the existing status of direct applications in current practice is intrinsically useful. From this starting point, the question arose of how more direct applications can occur. Those working in the field tend to express desires or expectations, for example, 'It is hoped that DDL for young EFL learners will become mainstream' (Hirata 2020: 103), but it is important for corpus linguists to reflect on how this can genuinely become possible. This research project's second contribution to the field has been to focus on this issue, and to suggest a practical solution. The solution offers differs from others in that it is presented to teachers as a pedagogical toolbox, based on a particular curriculum, rather than as corpus linguistics per se.

These contributions to the field are made within the scope of the research project, which is necessarily limited (6.6.3), but is a starting point for future directions of research (6.6.4).

6.6.3 The limitations of the research project

This research project is particular to one country's education system. There may be differences in other countries' curricula that make direct applications more difficult, or easier. There also may be differences between countries in the relationship between curriculum and classroom. Yet this dissertation's suggestion for addressing the lack of direct applications, taken at its broadest, can be applied in any geographical context. It would be a case of designing exercises and other materials to work with that country's classroom needs and curriculum. Limiting this research project's research question to Norway has been practical, in that it the research was undertaken in Norway, and useful, in that the research findings can benefit the English subject in Norway.

The questionnaire and interviews were of comparatively few teachers, so the data can be taken as an indication of teachers' opinions only. Yet this data provided qualitative insights into at least some teacher opinions on direct applications, which was enough to enable this research project to suggest a solution to help boost direct applications.

The view was put forward (3.2.4; 6.2) that because there are explicit language topics included in the current curriculum, classroom practices should be different than under the previous curriculum. As stated at the start of this chapter (6.1), the curriculum is something teachers need

to know and base their teaching on. Yet changes in classroom practices are not empirically verified here, and that is a limitation.

The third phase of the research limited its focus to pupil-corpus interaction, but there were findings on teacher-corpus interaction in an earlier phase (the first dissertation article, Kavanagh 2021a: 13-15). There has not been the available scope to explore solutions for teachers who would use corpora for reference and for preparing teaching, but see 6.6.4 below.

The analysis of textbooks in this research was limited. Only three textbook series were analysed. The reason was that the textbooks of three prominent publishers seemed enough to facilitate an understanding and overview of lower secondary English language exercises, but this cannot be known for certain. The approach has seemed reasonable, but the limitation can be noted.

As mentioned in the previous paragraph, the textbooks were for lower secondary school only. It is a necessary limitation of the project that there was no capacity to undertake the same work for primary and upper secondary school, but see 6.6.4 below.

As mentioned in the third dissertation article (footnote 5), there are also digital resources for the English subject for teachers in Norway, but the present author was not granted access to them. Local government provides these digital resources, and an investigation of language exercises in them (if there are any) could have influenced the design of the suggested corpus exercises. As it stands, the textbooks seem to have sufficed, but see 6.6.4 below.

The exercises related to vocabulary teaching and learning topics only. The reasons for this are given in the relevant article (the third dissertation article, section 4.2.1), but as a result, the exercises comprise a limited number of examples of what can be designed for teachers and pupils. The exercises are few in number, and cover only two vocabulary topics: synonyms and collocations. A small number of example suggested exercises is a modest design. However, this was a practical limit for the scope of a research project such as this one. The examples have been suggested as the springboard for a greater endeavour (6.5).

The suggested exercises have been presented to some teachers (see 4.3.2), but they have not yet been published. Thus, they have not been widely disseminated to the in-service teachers who might use them, and this means they have not been put to the test in the classroom. However, the exercises are for online corpora that have already been used with pupils (SKELL and Netspeak), and the classroom use of the exercises is perhaps something that could be observed in future research (see 6.6.4).

6.6.4 Future directions for research

Three future directions for research will be identified in this section. These are not mutually exclusive research directions, and they could be combined.

In the previous section, limitations of the present research were delineated. The first direction for future research proposed here is a continuation of the present research. The existing exercises could be tested on teachers and pupils. The exercises could also be expanded beyond lower secondary. Exercises for primary and upper secondary could be designed using the same methods. Digital resources for teachers (which the present author was not given access to) could be investigated to see if they could be enhanced or replaced by corpus exercises. Also, the collection of exercises could be expanded beyond vocabulary to grammar and phonology, and beyond language topics to culture and intercultural communication. The exercises used only two corpora – the use of SKELL and Netspeak emerged from the research data – but there will be other corpora appropriate for direct applications that could be utilized in an expansion of the collection. Research could also expand beyond pupil-corpus interaction to teacher-corpus interaction. Corpus-using teachers in the first phase of the research used corpora for reference and for preparing teaching, and research could be undertaken into how best to offer solutions for these activities for teachers who do not currently use corpora.

The second direction for future research proposed here is to develop ideas of collaboration and dialogue between corpus linguists, teachers and materials developers, as described already in section 6.4.4. To collaborate on pedagogical and curricular corpus exercises means one person is not required to have multiple skill sets in order to produce material. It is not likely that one person would be a corpus linguist, a teacher and a materials developer all at once. To successfully integrate corpora into the real-life classroom, these three areas of knowledge could be brought together. To research how this could be done effectively would be useful. Exercises and teaching materials should fit not only the curriculum, but other teacher needs and classroom foci, of which non-teachers may not be aware. A collaboration cannot be the kind of working relationship that gives rise to the ownership problem discussed above (6.4.4; 6.5), so research projects undertaken between these three areas of expertise need to carefully balance the various roles of the individuals, but such research could result in the design of very effective material.

The third direction for future research pertains to a finding from the first two dissertation articles that has not been hitherto raised in this discussion: the lack of interest of some pupils in language (Kavanagh 2021a: 15; Kavanagh 2021b: 100). This is worthy of further investigation. One body of theory and research that could be utilized is that of learning styles. A style is distinct from an ability: one style is not better than another. Learners have different styles of learning, or at least

use different learning styles to varying degrees, and there have been various theories of styles: cognition-centred, personality-centred and activity-centred (Sternberg & Grigorenko 2001). This is an area of research that could be brought into the design of direct applications. It may be that corpora are not compatible with all learning styles, or that corpus exercises need to be adaptable for different kinds of learning styles. For example, using a theory of Kolb's, deductive exercises may suit 'convergers', while inductive exercises may suit 'assimilators' (ibid.: 16-17). Research in this area can form part of a connection between corpus linguistics and the field of second language acquisition, which has often been called for (see O'Keeffe 2021). This is because theories of learning styles are engaged with in second language acquisition literature (see Ortega 2009: 192-215).

6.6.5 Returning to the 'bridge' between corpus linguistics and teaching practice

In Chapter 1 (1.4), it was stated that researchers have found there to be a 'gap' between corpus linguistics and teaching practice. The long-standing metaphor that has been used for the resolution of this is 'bridging the gap'. Suggestions for bridging this gap have included educating teachers about, or in, corpus linguistics; designing pedagogically motivated corpora; and improving the usability of software (see Chapter 3). As things stood at the outset of this research project, these had not yet resulted in increased direct applications. This research project has been able to indicate through its findings that there are few direct applications in Norwegian schools. It has focused on the pedagogical and curricular context, through data gathered from English teachers. This dissertation puts forward the idea that it might be more fruitful to 'bridge the gap' by starting from the teacher's 'side' of that gap. The suggested solution here is not to attempt to change the teachers (turn them into linguists in order for them to better comprehend corpora), but to attempt to show them how corpora can be useful, by suggesting how direct applications can easily be integrated into their work. Given enough time, perhaps all teachers could become the corpus linguists we would like them to be (see 6.4.1), but another approach can be taken in the meantime. In order to cross this 'bridge' now, we do not have to wait for better corpora, better software, or linguist-teachers to appear. Corpora have always been associated with language teaching and learning, and if language teaching and learning can be the starting perspective, it will be seen that there are aspects of corpus linguistics that can fulfil essential and relevant needs already. The suggested exercises are a step in this direction. The aim for direct pedagogical corpus applications can be to provide teachers with solutions for what is currently required in their English teaching: authentic language data, explicit language learning, language awareness, critical thinking, in-depth learning, and digital skills.

References

Allan, R. (2023), 'Reserved for Research? Normalising Corpus Use for School Teachers', *Nordic Journal of English Studies*, 22/1: 68-92.

Angelsen, A. and Hauge, H. (2020), 'Developing language awareness', in C. Carlsen, M. Dypedahl, and S.H. Iversen (eds.), *Teaching and Learning English* (2nd edn., Cappelen Damm Akademisk; Oslo), 322-342.

Anthony, L. (2022), AntConc (Version 4.2.0) [Computer Software] (Waseda University; Tokyo) <https://www.laurenceanthony.net/software>.

Baisa, V., and Suchomel, V. (2014), 'SKELL—web interface for English language learning', in A. Horák and P. Rychlý (eds.), *Proceedings of RASLAN 2014 (recent advances in Slavonic natural language processing)* (NLP Consulting, Czech Republic), 63–70. Details of the English Corpus for SKELL: <https://www.sketchengine.eu/english-skell-corpus/>, last accessed 28 August 2023.

Baker, P., Hardie, A., and McEnery, T. (2006), *A Glossary of Corpus Linguistics* (Edinburgh University Press; Edinburgh).

Bennett, S., Maton, K., and Kervin, L. (2008), 'The 'digital natives' debate: A critical review of the evidence', *British Journal of Educational Technology*, 39/5: 775–786. doi:10.1111/j.1467-8535.2007.00793.x.

Biber, D., Conrad, S. and Reppen, R. (1998), *Corpus Linguistics: Investigating Language Structure and Use* (Cambridge University Press; Cambridge).

Borg, S. (2015), *Teacher Cognition and Language Education: Research and Practice* (Bloomsbury Publishing; London).

Boulton, A. (2011), 'Data-driven learning: The Perpetual Enigma', in S. Goźdź-Roszkowski (ed.), *Explorations across languages and corpora* (Peter Lang; Frankfurt), 563-580.

Boulton, A., and Cobb, T. (2017), 'Corpus use in language learning: A meta-analysis', *Language*

Learning, 67/2: 348-393.

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.

Breyer, Y. (2006), 'My Concordancer: Tailor-made software for language learners and teachers', in S. Braun, K. Kohn, and J. Mukherjee (eds.), *Corpus Technology and Language pedagogy: New Resources, New Tools, New Methods* (Peter Lang; Frankfurt), 157-176.

Breyer, Y. (2009), 'Learning and teaching with corpora: reflections by student teachers', *CALL*, 22/2: 153-172.

Breyer, Y. (2011), *Corpora in Language Teaching and Learning: Potential, Evaluation, Challenges* (Peter Lang GmbH Internationaler Verlag der Wissenschaften; Berlin) <https://www.researchgate.net/publication/264499058_Corpora_in_Language_Teaching_and_Learning_Potential_Evaluation_Challenges>, last accessed 28 August 2023.

Brinkmann, S. and Kvale, S. (2015), *InterViews: Learning the Craft of Qualitative Research Interviewing* (Sage; Los Angeles, London, New Delhi, Singapore, Washington D.C.).

Buendgens-Kosten, J. (2013), 'Authenticity in CALL: Three Domains of "Realness"', *The Journal of EUROCALL*, 25/2: 272-285.

Buendgens-Kosten, J. (2014), 'Key concepts in ELT: Authenticity', *ELT Journal*, 68/4: 457-459.

Bullen, M., Morgan, T., and Qayyum, A. (2011), 'Digital Learners in Higher Education: Generation is Not the Issue', *Canadian Journal of Learning Technology*, 37/1.

Burner, T. (2020), 'What the 2020 curriculum reform means for English teachers', in C. Carlsen, M. Dypedahl, and S.H. Iversen (eds.), *Teaching and Learning English* (2nd edn., Cappelen Damm Akademisk; Oslo), 53-61.

Callies, M. (2019), 'Integrating corpus literacy into language teacher education: the case of learner corpora', in S. Götz and J. Mukherjee (eds.), *Learner Corpora and Language Teaching* (John

Benjamins Publishing Company; Amsterdam), 245-263.

Cardona, M.D., Didriksen, A.A., and Gjesdal, A.M. (2014), 'Korpusbasert undervisning i fremmedspråkene: la elevens nysgjerrighet sette dagsorden', *Acta Didactica Norge*, 8/2: 1-26.

Chambers, A. (2019), 'Towards the corpus revolution? Bridging the research-practice gap', *Language Teaching*, 52: 460-475. doi:10.1017/SO261444819000089

Charles, M. (2018), 'Use of corpus tools in writing and editing of master's theses and doctoral theses', seminar at Inland Norway University of Applied Sciences, 8-9 May 2018.

Cohen, L., Manion, L., and Morrison K. (2011), *Research Methods in Education* (Routledge; London and New York).

Cook, G. (1997), 'Language play, language learning', *ELT Journal*, 51/3: 224-231.

Council of Europe (2001), *Common European framework of reference for languages: Learning, teaching and assessment* <https://rm.coe.int/1680459f97>, last accessed 28 August 2023.

Council of Europe (2018), *Common European framework of reference for languages: Learning, teaching and assessment. Companion volume with new descriptors* <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>, last accessed 28 August 2023.

Crosthwaite, P. (ed.). (2020), *Data-driven learning for the next generation: Corpora and DDL for Pre-tertiary learners* (Routledge; London).

Crosthwaite, P. (2022), 'DDL for Younger Learners', in R.R. Jablonkai and E. Csomay (eds.), *The Routledge Handbook of Corpora and English Language Teaching and Learning* (Routledge; London and New York), 377-389.

Davies, M. (2002-), English-Corpora.org <https://www.english-corpora.org/>.

Davies, M. (2008-), Corpus of Contemporary American English (COCA): <https://www.english-corpora.org/coca/>.

Davies, M. (2013), Corpus of Global Web-Based English <https://www.english-corpora.org/glowbe/>.

Dekeyser, R. (2005), 'Implicit and Explicit Learning', in C.J. Doughty and M.H. Long, *The Handbook of Second Language Acquisition* (Blackwell; Oxford), 241-267.

Diskin, E. and Winsvold, K.G. (2021), *Enter 10 Engelsk* (Gyldendal; Oslo).

Dypedahl, M. and Hasselgård, H. (2018), *Introducing English Grammar* (3rd edn., Fagbokforlaget; Bergen).

Ebrahimi, A., and Faghih, E. (2016), 'Integrating corpus linguistics into online language teacher education programs', *ReCALL*, 29/1: 120–135. doi:10.1017/S0958344016000070.

Ellis N.C. and Robinson P. (2008), 'An introduction to cognitive linguistics, second language acquisition and language instruction', in P. Robinson and N.C. Ellis (eds.), *Handbook of Cognitive Linguistics and Second Language Acquisition* (Routledge; New York), 3-24.

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. doi:10.1080/09571736.2010.520728.

Farr, F. and Karlsen, P.H. (2022), 'DDL Pedagogy, Participants, and Perspectives', in R.R. Jablonkai and E. Csomay (eds.), *The Routledge Handbook of Corpora and English Language Teaching and Learning* (Routledge; London and New York), 329-343.

Fenner, A. (2020), 'The Historical Development of English as a School Subject', in A. Fenner and A.S. Skulstad (eds.), *Teaching English in the 21st Century* (2nd edn., Fagbokforlaget; Bergen), 17-42.

Fenner, A. and Ørevik, S. (2020), 'Analysis of learning materials', in A. Fenner and A.S. Skulstad, (eds.), *Teaching English in the 21st Century* (2nd edn., Fagbokforlaget; Bergen), 337-364.

Flognfeldt, M., and Lund, R. (2021), *English for Teachers and Learners* (2nd edition, Cappelen Damm Akademisk; Oslo).

Foddy, W. (1994), *Constructing Questions for Interviews and Questionnaires: theory and practice in social research* (Cambridge University Press; Cambridge).

Friginal, E. (2018), *Corpus Linguistics for English Teachers* (Routledge; New York and London).

Gaskell, D., and Cobb T. (2004), 'Can learners use concordance feedback for writing errors?', *System*, 32/3: 301-309. doi:10.1016./j.system.2204.04.001.

Gilje, Ø., Ingulfsen, L., Dolonen, J.A., Furberg, A., Rasmussen, I., Kluge, A., Knain, E., Mørch, A., Naalsund, M., and Skarpaas, K.G. (2016), *Med ARK&APP – Bruk av læremidler og ressurser for læring på tvers av arbeidsformer* (University of Oslo; Oslo).

Gilmore, A. (2004), 'A comparison of textbook and authentic interactions', *ETL Journal*, 58/4: 363-374.

Gilmore, A. (2007), 'Authentic Materials and Authenticity in Foreign Language Learning', *Language Teaching: The International Research Resource for Language Professionals*, 40/2: 97-118.

Gilquin, G. and Granger, S. (2022), 'Using data-driven learning in language teaching', in A. O'Keeffe and M.J. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics* (2nd edn., Routledge; London), 430-442.

Hasselgård, H. (2018), 'Corpora in English Language Teaching', in C. Bøhn, M. Dypedahl, and G. Myklevold, *Teaching and Learning English* (1st edn., Cappelen Damm Akademisk; Oslo), 322-342.

Hasselgård, H. (2020), 'Corpora in English Language Teaching', in C. Bøhn, M. Dypedahl, and S.H. Iversen (eds.), *Teaching and Learning English* (2nd edn., Cappelen Damm Akademisk; Oslo), online chapter.

Henry, C. and Sheepy H. (2022), 'Evaluating Corpus Analysis Tools for the Classroom' in R.R. Jablonkai and E. Csomay (eds.), *The Routledge Handbook of Corpora and English Language Teaching and Learning* (Routledge; London and New York), 437-459.

Heyvaert, L., and Laffut, A. (2008), 'Corpora in the teaching of English in Flemish secondary schools: Current situation and future perspectives', in A. Frankenberg-Garcia, T. Rkibi, M. R. Cruz, R. Carvalho, C. Direito, and D. Santos-Rosa (eds.), *Proceedings of TaLC 8—Lisbon, 8th*

Teaching and language corpora conference (Associação de Estudos e de Investigação do ISLA-Lisboa; Lisbon), 400–409.

Hirata, E. (2020), in P. Crosthwaite (ed.), *Data-driven Learning for the Next Generation* (Routledge; London and New York), 88-105.

Johns, T. (1991), 'Should You Be Persuaded – Two Samples of Data-driven Learning Materials', *ELR Journal Classroom Concordancing*, 4: 1-16.

Karlsen, P.H. (in preparation), 'Integrating Multimedia Corpora in the Secondary School Classroom in Norway'.

Karlsen, P.H., and 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.

Kavanagh, B. (2021a), '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.

Kavanagh, B. (2021b), 'Norwegian In-service Teachers' Perspectives on Language Corpora in Teaching English', *Nordic Journal of Language Teaching and Learning*, 9/2: 90-106.

Koellner, K, and Greenblatt, D. (2018), 'Inservice Teacher Education' entry in Oxford Bibliographies <https://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0196.xml>, last accessed 28 August 2023.

Leech, G.N. (1997), 'Teaching and language corpora: A convergence', in A. Wichmann, S. Fligelstone, T. McEnery, and G. Knowles (eds.), *Teaching and language corpora* (Longman; London), 1-23.

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. doi:10.1017/S095834401400010X.

Leńko-Szymańska, A., and Boulton, A. (2015), 'Data-driven learning in language pedagogy', in A.

Leńko-Szymańska and A. Boulton (eds.), *Multiple Affordances of Language Corpora for Data-driven Learning* (John Benjamins Publishing Company, Amsterdam and Philadelphia), ix-xxii.

Liu, D., and Lei, L. (2017), *Using Corpora for Language Learning and Teaching* (Tesol Press; Alexandria).

Lund, R. (2020), 'Teaching and learning materials', in C. Bøhn, M. Dypedahl, and S.H. Iversen (eds.), *Teaching and Learning English* (2nd edn., Cappelen Damm Akademisk; Oslo), 343-362.

McDonough, J., Shaw C., and Masuhara, H. (2012), *Materials and Methods in ELT: A Teacher's Guide* (3rd edn., Wiley-Blackwell; West Sussex).

McEnery, T., and 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)* (Routledge: London and New York), 364-380.

Madsen, T. and Mohammed-Roe, S. (2020), *Engelsk 8 for Cappelen Damm Student's Book* (Cappelen Damm; Oslo).

Meunier, F. (2022), 'Revamping DDL: Affordances of Digital Technology', in R.R. Jablonkai and E. Csomay (eds.), *The Routledge Handbook of Corpora and English Language Teaching and Learning* (Routledge; London and New York), 344-360.

Millar, N., and Lehtinen, B. (2008), 'DIY Local Learner Corpora: Bridging Gaps Between Theory and Practice', *The JALT CALL Journal*, 4/2: 61–72.

Morrow, K. (1977), 'Authentic texts and ESP', in S. Holden (ed.), *English for Specific Purposes* (Modern English Publications; London), 13–17.

Mukherjee, J. (2004), 'Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany', in U. Connor and T.A. Upton (eds.), *Applied corpus Linguistics: A Multidimensional perspective* (Rodopi; Amsterdam and New York), 239-250.

National Research Council. (2000), *How people learn. Brain, mind, experience and school* (Expanded edn., The National Academies Press; Washington D.C.).

Newby, D. (2020), 'Teaching Grammar', in A. Fenner and A.S. Skulstad (eds.), *Teaching English in the 21st Century: Central Issues in English Didactics* (2nd edn., Fagbokforlaget; Bergen), 213-238.

Nielsen, J. (1993), *Usability Engineering* (AP Professional; Boston).

Norges Offentlige Utredninger. (2015), *NOU 2015: 8. Fremtidens Skole – Fornyelse av fag og kompetanser* <https://www.regjeringen.no/no/dokumenter/nou-2015-8/id2417001>, last accessed 4 July 2022.

Norwegian Directorate for Education and Training. (2006), *English Subject Curriculum.*

Norwegian Directorate for Education and Training. (2013), *English Subject Curriculum* <https://data.udir.no/kl06/ENG1-03.pdf?lang=eng>, last accessed 23 September 2022.

Norwegian Directorate for Education and Training. (2017), *Core curriculum – values and principles for primary and secondary education* <https://www.udir.no/lk20/overordnet-del/?lang=eng>, last accessed 4 July 2022.

Norwegian Directorate for Education and Training. (2019), *Curriculum in English* <https://data.udir.no/kl06/v201906/laereplaner-lk20/ENG01-04.pdf?lang=eng>, last accessed 4 July 2022.

Norwegian Ministry of Education and Research. (2016), *Report to Parliament 28 (2015-2016) Subject – Specialization – Understanding. A renewal of the knowledge promotion* <https://www.regjeringen.no/no/dokumenter/meld.-st.-28-20152016/id2483955/sec1>, last accessed 20 December 2022.

Norwegian Ministry of Education and Research. (2022), *Læringsanalyse – noen sentrale dilemmaer: Delrapport fra ekspertgruppen for digital læringsanalyse* <https://www.regjeringen.no/no/dokumenter/laringsanalyse-noen-sentrale-dilemmaer/id2916747/>, last accessed 28 August 2023.

O'Keeffe, A. (2021), 'Data-driven learning: a call for a broader research gaze', *Language Teaching*, 54: 259-272. doi:10.1017/S0261444820000245.

Ortega, L. (2009), *Understanding Second Language Acquisition* (Hodder Education; UK).

Palma, T.F. (2022), 'Corpora for teaching culture and intercultural communication', in R.R. Jablonkai and E. Csomay (eds.), *The Routledge Handbook of Corpora and English Language Teaching and Learning* (Routledge; London and New York), 116-130.

Pellegrino, J.W. and Hilton, M. L. (2012), (eds.), *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century* (The National Academies Press; Washington D.C.).

Pérez-Paredes, P. (2020), 'The Pedagogic Advantage of Teenage Corpora for Secondary School Learners', in Crosthwaite (ed.), *Data-driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners* (Routledge; London and New York), 67-87.

Pettersen, S., and Røkaas, F. (2021), *Stages 10 Engelsk for ungdomstrinnet* (Aschehoug Undervisning; Oslo).

Pinto, P.T., Crosthwaite, P., de Carvalho, C.T., Spinelli, F., Serpa, T., Garcia, W., and Ottaiano, A.O. (2023) (eds.), *Using Language Data to Learn About Language: A Teachers' Guide to Classroom Corpus Use* (University of Queensland; Brisbane). doi:10.14264/3bbe92d.

Rødnes, K.A., and de Lange, T. (2012), *Læreres bruk av og erfaringer med veiledninger til læreplaner for fag og veiledning i lokalt arbeid med læreplaner* (University of Oslo; Oslo) <https://www.udir.no/globalassets/filer/tall-og-forskning/forskningsrapporter/lareres-bruk-av-og-erfaringer-med-veiledninger-til-lareplaner-for-fag-og-veiledning-i-lokalt-arbeid-med-lareplaner.pdf >, last accessed 28 August 2023.

Røkenes, F.M., and R.J. Krumsvik. (2016), 'Prepared to teach ESL with ICT? A study of digital competence in Norwegian teacher education', *Computers and Education* 97: 1-20 <http://dx.doi.org/10.1016/j.compedu.2016.02.014>, last accessed 28 August 2023.

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* (John Benjamins Publishing Company; Amsterdam), 83-98.

Römer, U. (2011), 'Corpus research applications in second language teaching', *Annual Review of Applied Linguistics* 31: 205-225. doi:10.1017/S0267190511000055.

Sawyer, R.K. (2006), 'The New Science of Learning', in R.K. Sawyer (ed.), *The Cambridge Handbook of The Learning Sciences* (Cambridge University Press; Cambridge).

Schaeffer-Lacroix, E. (2020). 'Barriers to trainee teachers' corpus use', in P. Crosthwaite (ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners* (Routledge; London and New York), 47–64.

Schulmeister, R. (2009), 'Is There a Net Gener in the House? Dispelling a Mystification', *eleed*, 5 <https://www.eleed.de/archive/5/1587>, last accessed 28 August 2023.

Sealey, A., and Thompson, P. (2007), 'Corpus, concordance, classification: Young learners in the L1 classroom', *Language Awareness*, 16/3: 208-223.

Simensen, A.M. (2020), 'A brief history of English teaching and learning in Norway', in C. Carlsen, M. Dypedahl, and S.H. Iversen (eds.), *Teaching and Learning English* (2nd edn., Cappelen Damm Akademisk; Oslo), 19-39.

Skjervheim, H. (1957, 1996), *Deltakar og tilskodar og andre essays* (Aschehoug, Oslo).

Skulstad, A.S. (2020), 'Communicative competence', in A. Fenner and A.S. Skulstad (eds.), T*eaching English in the 21st Century: Central Issues in English Didactics* (2nd edn., Fagbokforlaget; Bergen), 43-68.

Speitz, H. (2020), 'National curricula and international perspectives', in C. Carlsen, M. Dypedahl, and S.H. Iversen (eds.), *Teaching and Learning English* (2nd edn., Cappelen Damm Akademisk; Oslo), 40-52.

Sternberg, R.J., and Grigorenko, E.L. (2001), 'A Capsule History of Theory and Research on Styles', in R.J. Sternberg and L. Zhang (eds.), *Perspectives on Thinking, Learning, and Cognitive Styles* (Lawrence Erlbaum Associates Publishers; London).

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.

Tomlinson, B. (2012), 'Materials development for language learning and teaching' *Lang. Teach.*, 45/2: 143–179. doi:10.1017/S0261444811000528.

Vitaz, M. and Poletanović, M. (2020), 'Data-driven Learning: The Serbian Case', *EL.LE,* 8/2: 409-422.

Vyatkina, N. (2016), 'Data-driven Learning of Collocations: Learner Performance, Proficiency, and Perceptions', *Language Learning & Technology*, 20/3: 159-179.

Wray, A. and Bloomer, A. (2006), *Projects in Linguistics* (2nd edn., Hodder Education; London).

Xu, J. (2022), 'A Historical Overview of Using Corpora in English Language Teaching', in R.R. Jablonkai and E. Csomay (eds.), *The Routledge Handbook of Corpora and English Language Teaching and Learning* (Routledge; London and New York), 11-25.

Yoon, H. and Jo, H. (2014), 'Direct and Indirect Access to Corpora: An Exploratory Case Study Comparing Students' Error Correction and Learning Strategy Use in L2 Writing', *Language Learning & Technology*, 18/1: 96-117.

Zareva, A. (2017), 'Incorporating corpus literacy skills into TESOL teacher training', *ELT Journal*, 71/1: 69-79. doi:10.1093/elt/ccw045.

Dissertation articles

1

Errata

Page	Published text	Correct text
3	it one of only 12 countries to have	it is one of only 12 countries to have
29	No collocation	Collocation

Bridging the Gap from the Other Side: How Corpora Are Used by English Teachers in Norwegian Schools

Barry Kavanagh, Inland Norway University of Applied Sciences

Abstract
Researchers have written of 'bridging the gap' between corpus linguistics and teaching practice. This study focuses on in-service English teacher informants from Norwegian schools, to try to address the 'gap' from the teaching practice 'side', rather than from the linguist 'side' engaged in spreading corpus linguistics. The study collects data on teachers' familiarity with corpus linguistics, what corpora are used for and how, and teachers' views on the obstacles to corpus use. The research question is *How are corpora used by in-service English teachers in Norwegian schools?* The research design consists of an online questionnaire and follow-up interviews. The questionnaire was answered by 210 teachers, 34 of whom answered they had done some work with corpora. The interviews were with three corpus-using teachers. The corpora they used were GloWbE, SkELL, Netspeak and COCA. Teacher-corpus interaction was for reference and for creating vocabulary and varieties of English exercises, and pupil-corpus interaction was encouraged by two of the teachers. The obstacles to the use of corpora were identified as differences between school levels, usability, and lack of teacher need. In concluding remarks, it is suggested that a starting point for corpus use among teachers may be to teach the tools and methods that seem to be already working for in-service teachers.

Keywords: corpus; corpus linguistics; English language teaching; in-service teachers; Norwegian schools

1. Introduction

There are ways in which corpora can be pedagogically valuable in language education. To list some: the authentic language found in corpora can show learners how a language is actively used; corpora are richer sources of data about a language than traditional reference books; and learners involved in a data-driven learning process can, through corpora, notice linguistic features, an activity which relates to the 'noticing' hypothesis of second language acquisition, and to learning motivation theory (Lin & Lee 2015: 264-5).

These declared potential benefits apply to language learning in general, but this article has a more specific scope, the role of corpora in language teaching in Norwegian schools (at all levels, i.e. both primary and secondary). There has been research into the potential of using

Kavanagh, Barry. 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.

corpora in schools (e.g. Braun 2007; Crosthwaite 2020), but it has been acknowledged that in Norway, corpora have not been used in language teaching to any great degree (Cardona, Didriksen & Gjesdal 2014: 1). It has been suggested, but not empirically established, that 'lack of time, group sizes and technological obstacles can stand in the way' (Ibid.). This apparent low level of direct applications[1] of corpora in schools mirrors the international picture, where researchers have found such a low level, and have discussed the difficulty of 'bridging the gap' (Mukherjee 2004; Breyer 2011: 146) between corpus linguistics and teaching practice. The difficulties identified have included not enough teachers 'systematically familiarized with the basic foundations, implications and applications of corpus linguistics' (Mukherjee 2004: 243) and the lack of '(classroom) userfriendly concordancing software' (Breyer 2011: 207).

This study aims to discover whether and what corpus linguistics tools and methods are being used in Norway by in-service teachers in their teaching. In-service 'designates 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). A country-wide survey of how familiar in-service English teachers are with corpus linguistics has been undertaken for this study, and from the survey, corpus-using teachers have been identified. Informants are included irrespective of their corpus literacy,[2] firstly because the intention is to gather data from as many teachers as possible, and secondly there may be obstacles to corpus use that are not affected by the corpus literacy of the teacher.

Some previous research in Germany has surveyed in-service teachers about their corpus use (Mukherjee 2004; Callies 2019), the results of which are discussed in section 2. The present study has gathered data on how corpus-using teacher informants use corpora, what they use them for, and what they find the obstacles to corpus use to be. This is to try to address the abovementioned difficult 'gap' from the teaching practice 'side', rather than from the linguist 'side' engaged in spreading corpus linguistics, where there is already a body of research on corpus

[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: 207).
[2] For a detailed definition of corpus literacy see Callies 2019: 247.

instruction for pre-service language schoolteachers (Breyer 2009 & 2011; Farr 2008; Heather & Helt 2012; Leńko-Szymańska 2014; Zareva 2017). To collect data from teachers who are already in service, and using corpora in their teaching, is potentially illuminating.[3] It may become a starting point for spreading the use of corpus linguistics among teachers, so investigating the school context may be insightful for the corpus education of teachers, a 'bridge' from teaching practice to corpus linguist. The research question is *How are corpora used by in-service English teachers in Norwegian schools?*

This specifies English teachers, and the English subject in Norway is a fitting one for conceivable corpus use, because there are potential obstacles that are not present. First, the English subject curriculum presents no obstacle, in principle, to the use of corpus methods. Although corpora are not mentioned in it, the curriculum has competence aims[4] that could be fulfilled by corpora, namely aims related to patterns in the language, to language use in context, to English and one's own language(s), and to to digital resources or tools (Utdanningsdirektoret 2013, 2019). Second, the Norwegian education system has commensurate IT infrastructure; it is the norm for teachers and pupils to have access to computer equipment and an internet connection. Third, Norway is not a country that has low proficiency in English; it one of only 12 countries to have 'very high proficiency' status in Education First's *English Proficiency Index*, ranking third in the world (Education First 2019: 6).

Previous research is considered in the next section (2), specifically studies concerning in-service teachers' knowledge of corpora, and concerning what corpus-using teachers do with corpora. Then, the methods of this study are explained (section 3). The results are presented in section 4. In the final section, the results are discussed, with some concluding remarks.

2. Previous Research

The main focus of this study is on in-service teachers in schools, therefore this section is devoted to studies which have that focus.

[3] Data from non-users to find out why they do not use corpora, although worth collecting, is not within the scope of this study.

[4] 'Competence aims' is the term used in the curriculum.

(Teacher educators and higher education teachers work with adult language learners in non-compulsory education and, for example, corpus use considered useful in the third level classroom (Lin & Lee 2015) might not apply to school pupils.)

The studies of Mukherjee (2004) and Callies (2019) investigated how familiar corpus linguistics is to in-service teachers in schools. Mukherjee used a questionnaire to survey 248 German teachers of English in secondary schools. He found that 79.4% of his in-service teachers, before taking a corpus workshop, agreed with the statement 'No, I don't know anything about corpus linguistics' (Mukherjee 2004: 241). He concluded that this 'illustrates the low extent to which corpus linguistics has so far had an impact on teaching practice in Germany' (Mukherjee 2004: 242). Fifteen years later, Callies also used a questionnaire to survey German teachers of English in secondary schools, and interpreted the findings from his small-scale survey of 26 teachers as reconfirming Mukherjee's (Callies 2019: 252): only 34.6% of his in-service teachers had 'heard of' corpus linguistics in their university studies, and only 3.8% in their in-service practical teacher training (Callies 2019: 250). Callies does, however, note an increased awareness among teachers of corpus linguistics between his study and Mukherjee's, and sees it as due to 'a younger generation of language teachers who have been trained in the use of corpora for research purposes in their university studies', but nevertheless he observes that 'this seems not to have made a significant impact on their teaching practice' (Callies 2019: 252). To these two studies can be added that of Zareva (2017), who surveyed 21 TESOL teachers taking a master's degree in the USA. Although her informants were pre-service teachers, there were in-service teachers mixed in with them. She issued a questionnaire at the conclusion of a grammar course that included a corpus component (Zareva 2017: 71-3), and found that 43% of her informants said they 'knew what a language corpus was' before their course began. That percentage, however, is reduced to 29% when counting only those who had 'done basic corpus searches before' (Zareva 2017: 74). So in all three studies, a minority had awareness of corpora, which indicates no wide use of them in English teaching.

Two of these studies, Callies (2019) and Zareva (2017), and a study by Farr (2008), investigated what teachers knowledgeable of corpus linguistics do with corpora. Of these, only Callies's informants were

exclusively in-service teachers, but the others are considered here because the informants were a mix of pre-service and in-service teachers. Farr used a questionnaire to survey 25 teachers taking an MA in ELT in Ireland, and found that a majority of her informants were willing to use corpora in preparation for teaching or teaching materials, to use corpora in class, and to 'initiate' pupils, as long as there were computers; her informants predicted they could use corpora in ways such as researching 'words before teaching them' and taking 'real, authentic examples of use', to give two examples (Farr 2008: 37-38). However, Farr could not obtain a full picture of what her informants went on to do in their teaching. Similarly, some of Zareva's informants 'pointed out specific areas of teaching where they can apply their corpus research knowledge, of instance in teaching grammar, academic vocabulary, and the use of words and collocates' (Zareva 2017: 75), but we do not know whether they subsequently used their corpus skills in teaching. This is a problem, acknowledged by both Farr and Zareva, when gathering data from pre-service informants. Farr wondered 'whether these good intentions are realised when the STs [pre-service teachers] hit the reality of their first teaching position' (Farr 2008: 39), and could not answer this conclusively; Zareva cited Farr on this point (Zareva 2017: 75). The corpus use of Callies's informants is actual not potential use. Most of his informants who used corpora did so as reference, 'for example when marking or checking language acceptability, while corpora are hardly, if at all, used in learner-centred activities or to compile local learner corpora' (Callies 2019: 252).

Our state of knowledge about in-service teachers using corpora in English teaching in schools is thus limited to survey results (and none from Norway) in which only a minority of teachers claim to know about corpora or corpus linguistics; and in which teachers who declare they use corpora for teaching mostly use them for reference. This low awareness and low level of use of corpora may be a picture of English teaching around the world, and there is a need for research to show, from more countries (e.g. Norway), whether there are corpus-using English teachers, what they do, and to what degree they involve pupils. The present study obtains data on this.

3. Methods

The research design consists of an online questionnaire and follow-up interviews. One aim of the questionnaire was to provide data about corpus use in English language teaching in schools in Norway, rather than rely solely on statements that there was not much use (e.g. Cardona, Didriksen & Gjesdal op. cit.). To this end, the questionnaire was released online so that it could reach all districts of Norway. The questionnaire also had the purpose of identifying a subgroup within the sample: corpus-using teachers, who could then become informants for follow-up interviews (how informants were selected is explained below in section 3.2).

3.1. The questionnaire

The questionnaire was designed on, and distributed through, the platform Qualtrics.com. As a tactic to ensure a high response rate, it was decided to make it answerable in a few minutes, therefore not many questions were asked (Cohen et al. 2018: 264). The questionnaire began with questions that collected nominal data: teaching level, geographical location, years of teaching experience and age. There followed questions about teaching activities which informants could possibly use corpora for: English grammar; English usage; and an overview of pupil mistakes (questions 5, 6 and 7, in Appendix 1), without these questions leading the informants by mentioning corpora. The focus shifted to corpus linguistics with question 9, shown in Figure 1 (it is similar to a question from Leńko-Szymańska 2014: 268). Answers divided the informants into four (potentially five) groups, for different follow-up questions based on what they claimed to know.

Q9 Have you ever heard the linguistic term corpus? Choose from one of the answers below.
o I have never heard the linguistic term corpus before.
o I have heard the linguistic term corpus before, but I have little or no idea what it is.
o I am fairly familiar with corpus linguistics, but I have never done any practical work with corpora.
o I have already done some work with corpora.
o If you have an answer that you are sure is different to the ones above, please give it here:

Figure 1. The first corpus question in the questionnaire (question 9), English-language version.

There were four optional answers to the question, and an additional open option selected by no informants, indicating that the four options sufficed. The question uses familiarity with gradation, from a relationship with corpora that involves having worked with them, to never having heard of them. This feature was influenced by previous questionnaires (Breyer 2011: 162; Leńko-Szymańska 2014: 268).[5]

Three questions (questions 10, 12 and 13 in Appendix 1), asked of every informant, checked informant claims of familiarity. This was a feature of a previous questionnaire (Leńko-Szymańska 2014) in which informants were asked about their prior knowledge of the term *corpus*: 61.5% claimed they had 'a rough idea what it is', and their answers were cross-referenced with answers to open questions asking them to define *corpus*, *concordance* and *concordancer*. Some 'inaccurate, vague or even meaningless' answers were received, indicating it was an effective way of checking claims (Leńko-Szymańska 2014: 269).

Question 10 asked whether informants knew each of the terms *collocation*, *colligation*, *concordance*, *concordancer*, *frequency list*, *key word in context (KWIC)*, and *part of speech (POS)*. Question 12 showed screenshots of online corpus interfaces, including Sketch Engine for Language Learning (SkELL) (Baisa & Suchomel 2014) and english-corpora.org, where among others the Corpus of Contemporary American English (COCA) can be found (Davies 2008-). The text of the question accompanying the screenshots asked informants whether they knew

[5] These questionnaires were for pre-service teachers.

'these web tools'; this general phrase was chosen so as not to lead informants by identifying corpora using the specific term. Question 13 asked if informants know other 'web tools' similar to the ones shown. It is evident that these checks on claims to familiarity themselves involve claims, but as mentioned above, the questionnaire was designed to be taken in minutes, which precludes onerous tests of knowledge.

There were two questions (questions 16 and 17 in Appendix 1) for those who selected the option in question 9, 'I have already done some work with corpora.'[6] Question 16 asked in what context these informants had used corpora. Of the six options, three were relevant to direct applications, namely 'I have used corpora to check acceptability of use when in doubt (or when marking),' 'I have used corpora-based materials in my teaching' and 'I have introduced corpora to pupils.' These options were developed from a previous questionnaire (Callies 2019: 251, Table 5). Question 17 asked what these informants used corpora for, options being grammar, vocabulary, spelling, idiom, authentic dialogue, learner language, and 'other'. These options were not intended to be mutually exclusive uses of corpora. They can be seen as relevant to direct applications when cross-referenced with the answers to question 16. As the questionnaire was designed to have few questions, there were no further questions about these uses. Instead, the follow-up interviews were used to gather more detail.

The Norwegian Centre for Research Data (NSD) approved both the questionnaire and the terms of service with Qualtrics. On the first screen, before informants commenced, they were given information on their rights and how data would be used. By commencing the questionnaire, they gave consent.

The questionnaire was distributed in March and April 2018. It was answerable by computer or phone. Informants had the option of answering in either English or Norwegian. The questionnaire was publicized through social media groups of English teachers in Norway. The total membership of these groups was over 10,900. The total population of English teachers in schools in Norway is not known.

The questionnaire was answered by 210 informants Although this a fraction of the membership of the social media groups, the number of

[6] This article focuses on corpus-using teachers, but follow-up questions were also asked of non-corpus-using informants (see Appendix 1).

informants is substantial compared with relevant past surveys of teachers described in section 2 (Mukherjee 2004; Callies 2019). Proportionally, the sample size, 210, is far greater in this study than the 248 of Mukherjee (2004), as Germany's population is 16 times greater than Norway's.

3.2. The interviews

Informants were selected for interview in the following way. Those who answered in question 9 they 'have already done some work with corpora' and subsequently in question 16 answered either 'I have used corpora-based materials in my teaching,' 'I have introduced corpora to pupils,' or 'Other', were asked at the end of the questionnaire whether they were willing to be interviewed. Those who volunteered their emails were contacted later, and three consented to face-to-face interviews. The interviews were approved by NSD. Each informant was given an information letter about their rights of participation, and the use of the data, and signed a consent form.

The interviews were conducted in October and November 2018. The researcher visited the teachers at their schools, and made audio recordings. Each interview was approximately an hour long, and in English. The researcher transcribed the recordings using Express Scribe for playback and Word for text. Any repetitions, pauses or verbal tics are not included when quoting from the transcripts, for reasons of clarity; content, not delivery, is the focus of interest. (The researcher accepts that interpretations of what constitutes repetition and so on are his own.) For each interview, the researcher created a guide containing a set of questions (Appendix 2). These were related to the informant's questionnaire answers. For instance, if the informant had answered in question 16, 'I have used corpora to check acceptability of use when in doubt (or when marking),' then in the interview they were asked how.[7] Each interview guide is therefore unique.

From guide questions it can be seen that for each aspect of language that informants in questionnaire question 17 selected that they used corpora for (e.g. vocabulary or grammar), the interview data can show which direct applications were teacher-corpus interaction and which were pupil-corpus interaction, for these informants.

[7] See the second question in the 'Joy' interview guide (Appendix 2).

Apart from the guide questions, the interviews were conducted in a back-and-forth conversational manner typical of semi-structured interviews. For this article, the anonymized informants have names assigned from the random name generator at behindthename.com/random: Joy, Lars, and Thomas.

Joy teaches at all levels of upper secondary school (Vg1, Vg2, and Vg3),[8] and had 25 years' experience teaching in the Norwegian school system at the time of the questionnaire. She discovered corpora by finding the Corpus of Global Web-based English (GloWbE) (Davies 2013) during a web search. Lars teaches in lower secondary (grade 9 at the time of the questionnaire; grade 10 at the time of the interview), and had four years' teaching experience at the time of the questionnaire; he studied corpus linguistics as part of his master's degree. Thomas teaches Vg1 and had three years' experience at the time of the questionnaire; he also studied corpus linguistics as part of his master's.

4. Results

In the response to the questionnaire, there was a spread of informants: they taught at all levels, with 64 teaching primary, 71 lower secondary, and 84 upper secondary (with some overlap); they were geographically spread, from 18 of the 19 Norwegian counties of the time;[9] their teaching experience ranged from less than 1 year to 50 years; and they were in all age categories from 20-29 to 60+.

4.1. Familiarity with corpus linguistics

To the questions about teaching activities which informants could possibly use corpora for, such as question 6 dealing with English usage (Appendix 1), none of the informants specified the use of corpora. This indicates that corpora were not to the forefront of informants' minds in relation to teaching, even if they were familiar with them.

---

[8] The Norwegian school system consists of primary school (*Barneskole*, grades 1-7, ages 6-13), lower secondary school (*Ungdomsskole*, grades 8-10, ages 13–16), and upper secondary school (*Videregående* skole, grades Vg1-Vg3, ages 16–19).

[9] Subsequent to this, Norway reduced the number of its counties to 11.

In the answer to question 8 (Ibid.), 157 out of 188 informants claimed to collect pupil texts. This indicates a potential interest in local learner corpora, from which teachers can find 'common and persistent errors' (Callies 2019: 253), and can be used to create materials (Millar & Lehtinen 2008).

Question 9 (Figure 1) was answered by 193 informants. Each option of the question showed a different 'grade' of familiarity:

29 informants 'never heard' (15%)
75 'little or no idea' (39%)
55 'fairly familiar' (28%)
34 'done some work' (18%)

These answers were cross-referenced against the answers to questions 10, 12 and 13 to check claims of familiarity. There were 34 informants who claimed 'I have already done some work with corpora,' and when compared with other answers, it could be seen that fewer than that are familiar with important terms in corpus linguistics. For example, only 26 claim to know what *concordance* means. Unfamiliarity with terms indicates that the 34 who claim experience should not all be thought of as all equally knowledgeable about corpora.

33 of the 34 answered question 16 about the context of their corpus use. Only 12 of the 33 claimed to have used corpora in either or both of the contexts 'I have used corpora-based materials in my teaching' and 'I have introduced corpora to pupils.'[10] Of these, 11 said they had used corpora-based materials in their teaching (including all 3 interviewees), and 6 said they introduced corpora to pupils (including Joy and Lars).

4.2. What corpora are used for and how

Of the abovementioned 12 informants, what they answered they used corpora for in question 17 were:

[10] There are contexts that could be included or excluded when considering what constitutes using corpora in teaching. If the context 'I have used corpora to check acceptability of use when in doubt (or when marking)' is added, the number of informants is 17 out of 33, but that activity could include non-teaching-related checking.

Vocabulary (11, including all three interviewees)
Idiom (10, including all three interviewees)
Authentic dialogue (8, including Lars and Thomas)
Grammar (6, including Joy and Lars)
Learner language (5, including Joy and Thomas)
Other: spelling (2, including Joy)
Other: language and gender (1)
Other: turn-taking (1)
Other: varieties of English (1, Joy only)

The three questionnaire informants who volunteered for interview came from this pool of 12. The corpora they used were all online corpus interfaces. They used no offline tools such as downloadable concordancers.

In the questionnaire, when asked about teaching activities (questions 6 and 8), none of the three mentioned corpora, even though they all have used corpora in their teaching. They were asked about this in the interviews, and each gave a different reason why not. Joy said she did not think of corpora as a separate category to dictionaries: 'For me it just functions as a different type of dictionary, really.' Lars said he did not mention corpora because he tends to think only of the classroom when questions are formulated with language like 'when you teach' or 'your teaching', and in the classroom, 'I seldom actually show my students [corpora], like, all of them together'. For Thomas, he simply did not mention corpora because he had stopped using them by the time of the questionnaire.[11] So there is no single reason corpora might not be mentioned; they might go unmentioned by people who have actually used them (as with Julia and Lars).

The most relevant data from the interviews is here divided into four subsections. The first concerns the corpora that are used. For the next two, the purposes that these teachers use corpora for can be divided into 'teacher-corpus interaction', when they use a corpus themselves, and 'pupil-corpus interaction', when the pupils conduct searches of a corpus (Römer 2011: 207). The final subsection concerns what informants do not use corpora for.

[11] That Thomas stopped does not mean his other questionnaire answers are innacurate. For example, options in question 16 begin 'I have used…' and 'I have introduced…' etc.

4.2.1. The corpora that are used[12]

GloWbE, described above, is used by Joy, and she recognized the english-corpora.org interface in the questionnaire. Joy also uses SkELL, and she recognized that interface too. Netspeak.org is a search engine of English native speaker text, created at Bauhaus-University Weimar, and uses Google Books as its corpus. It is used by Joy, and she mentioned it in her response to question 13. Lars uses COCA, and he recognized the english-corpora.org interface in the questionnaire. It was also possibly used by Thomas, who said he used 'COCA or something like it' in a teaching context: he recognized the english-corpora.org interface in the questionnaire, but was genuinely unsure which of the corpora he had used. All three interviewees also use online collocation dictionaries. There are at least three popular online collocation dictionaries, and Lars specified Ozdic. As these are dictionaries, when they are corpus-informed (this information is not always given), they can be considered indirect applications (see section 1, footnote 1), but not corpora.

4.2.2. Purposes corpora are used for: teacher-corpus interaction

Teacher-corpus interaction is shown in Table 1. The purposes shown are elaborated upon below.

Table 1. Teacher-corpus interaction.

Informant	Corpus	Purpose: vocabulary	Other purpose
Joy	SkELL and Netspeak	Collocation	
	GloWbE		Spelling and varieties of English

[12] A limitation of the questionnaire is that it did not collect data on what corpora were used by the 9 non-interviewed teachers who used corpora in teaching. This may have been caused by being optimistic about interviewing all 12. The questionnaire at least collected data on what corpus interfaces they recognized from ones shown to them (see Q12, Appendix 1).

Lars	english-corpora.org (COCA)	Collocation, idiom	Checking acceptability
Thomas	English-corpora.org	Frequency	

Joy uses SkELL and Netspeak for vocabulary. She uses them to check collocation herself, 'to help students figure whether collocations are good, yes or no', in feedback on pupils' written work, or feedback in the classroom. When the latter occurs Joy uses SkELL, not the pupils, so it is not pupil-corpus interaction.

Lars mainly uses COCA to check acceptability of use. He uses COCA, because '99% of my students speak and write American English.' He also uses it for vocabulary, in the following manner. First he checks COCA himself, then he tries 'to have my students guess, like what other words go with this word, and then come with all their suggestions, and then I say, "These are the 20 most common words to use with this word." And then sometimes you'll get 10 out of 20 right, and then we can talk about the rest, "Why is this a word that comes up all the time?"' He does not present COCA's concordance lines to them; he re-types the examples into a Powerpoint slide (e.g. he shared with the researcher a slide where the word was *choice.*) He does not use the term *collocation* with pupils, but rather asks, 'What other words go with this word?' (*No choice, your choice*, etc.). Finally, Lars checks COCA when confronted with Norwegian idiomatic expressions that he suspects do not translate into English. 'So, do you have a phrase [written by a pupil] and it feels idiomatic in Norwegian? Put it into COCA and see if you can get someone else using that… I try to actually check, if it's a use I just haven't seen before.'

Thomas's main use for an english-corpora.org corpus was vocabulary. He 'used a corpus in planning of classes' and when he 'needed to look up some examples and get numbers and statistics for them [i.e. pupils], to show this is a better way of saying it than this way, because this is what they [i.e. English speakers] use.' A corpus-based vocabulary exercise that he has given to pupils involved the 100 most frequent words of English, asking them to write something with these

words: 'You'd be surprised at how many of them actually manage to string together authentic speech.'

Joy occasionally uses GloWbE for spelling and varieties of English. For her, these two aspects of language are intertwined. For spelling, normally Joy 'would use a dictionary' if she was in doubt. 'But of course you could use it [a corpus] for checking differences [...] then we're basically into varieties of English.' In Vg1, she teaches 'a quick session on British versus American [English]… getting them to be aware of the differences and [to] try and be consistent'. She said that 'once or twice' she showed the pupils the results of a GloWbE search, but did not have them search that corpus themselves because she thought 'that's too advanced for students', and that the average pupil is not interested in language.

4.2.3. Purposes corpora are used for: pupil-corpus interaction

Pupil-corpus interaction is shown in Table 2. The purposes shown are elaborated upon below.

Table 2. Pupil-corpus interaction.

Informant	Corpus	Purpose: vocabulary	Purpose: grammar
Joy	Netspeak	Collocation, idiom	Verb forms
	SkELL	Idiom	
Lars	english-corpora.org (COCA)	Checking acceptability	Suffixes

For vocabulary, Joy finds Netspeak useful for pupils finding collocations themselves: 'I think I managed to get some of them to start using this.' For idiomatic expressions in English, Joy has suggested to pupils that they use Netspeak or SkELL. 'If there's an expression they're not sure about, find it hard to find, you could use that [a corpus] instead of a dictionary,' she said, but she does not know whether her pupils actually do.

For grammar, while Joy would 'usually refer them [pupils] to the grammar section' of a textbook, she has introduced Netspeak to pupils

for verb forms. 'They can fill in, for instance [...] *look forward to*. Typically Norwegians use an infinitive, so then I say, "Okay, you've got to write *look forward to* with your question mark."' In Netspeak, a question mark is used as a symbol for representing any letter or number (i.e. a 'wild card'); the most frequent collocations of *look forward to* in a Netspeak search are *seeing* and *hearing*, and this would reveal that the *-ing* verb form is used, not the infinitive.

For vocabulary, Lars has shown some grade 10 pupils how to use COCA, in written feedback. He said his best pupils do not 'tend to make [...] concord mistakes or grammar mistakes, it's often usage mistakes'. For example, a pupil wrote 'I slowly exposed my eyes' and he had the pupil search COCA. The pupil then 'agreed that it sounded off'.

For grammar, Lars also taught the highest-performing pupils how to use in COCA 'a wild card to see if you can put suffixes on a word'. He gave the example of adjectives, and the *–ly* suffix for creating adverbs out of them.

4.2.4. Non-use of corpora

None of the three informants considered investigating concordances of learner texts. Joy has collected six or seven years of pupil texts. Her use of learner language occurs when on occasion she gives pupils 'texts written by other students and have them grade it'. She has not used a concordancer to search for patterns in these texts: 'I don't think I have any need for that really [...] I know where they make their mistakes.' When that method was suggested to Thomas, he reacted differently: 'Once you actually put that idea in my head that actually sounds very interesting [...] you can start just basically looking up different N-grams and see what repeats.' All three interviewees had collected pupil texts, but apparently had not thought of their collections as learner corpora, or even corpora.

4.3. Obstacles to the use of corpora

The obstacles to the use of corpora in teaching identified by these three informants can be divided into subsections: differences between school levels; usability; and lack of teacher need. These obstacles apply to both teacher-corpus interaction and pupil-corpus interaction.

4.3.1. Differences between school levels

In upper secondary school, when most pupils have a higher level of English, and would also be, perhaps because of age, better able to understand corpus methods and tools, there is actually less focus on the English language itself in the English subject. The teacher cannot change this by becoming more corpus literate. Less focus on language means there is less need for language corpora. At that level there is more focus on being to able to communicate about culture, society and news items. Joy said, 'There are so many other things we have to teach we do not really get to go into language.' Thomas referred to 'the amount of work that we have to put into social studies, pretty much, like my students right now are working with the #MeToo movement, explaining the American election, especially at this level'. This is a combination of the pressure of time and the pressure of the syllabus on the teachers. A related point from Thomas is about assessment at this level. In an exam, pupils are not marked solely for their language use:

> With language only really being a third of the grade you give, when I read, I'm so lost in paragraph structure and argument structure, and 'Do these facts comply? Are these citations correct?' that when you come to the language, the ones who do have language mistakes to the level where this will impede their grade, it's usually concord, verb conjugation, and I don't need to look up a corpus to see an is/are mistake.

At lower levels, where there is more focus on the linguistic aspects of the subject, advanced concepts like collocation (for example) are not taught to pupils. There was the example above of Lars only showing COCA to his best-performing pupils. The more advanced language patterns that can be discovered in corpora are less relevant to pupils at lower levels. Lars also sees a disadvantage in using corpora for examples of authentic dialogue. He prefers to use movie dialogue with pupils, because he notices 'how chaotic actual conversation is often. And you may not want to introduce that to tenth graders'.

4.3.2. Usability ('user friendliness')

The issue of usability came up in the discussions around COCA. While Lars could show COCA to a high-performing grade 10 student, he agreed he 'probably could' do this in grades 8 or 9, but said, 'It's a little confusing, the COCA interface is kind of messy… There's a lot of

buttons to push and it's kind of slow.' Thomas echoed this when he explained why he does not use a corpus with pupils. He used a corpus in class with pupils only once but the pupils did not understand what they saw. He explained, 'Most corpora look like web 2.0, they're not very user-friendly.' While educators can improve the digital competence of pupils (Ståhl 2017) and teachers (Røkenes & Krumsvik 2016), this may have limits and it does not rule out changing the software as the solution (Breyer 2011: 93).

4.3.3. Lack of teacher need

This is an issue for teachers regardless of their level of corpus literacy. Thomas said he used to check corpora for himself when giving written pupil feedback, but stopped: 'I started using it, I started out as a teacher using it, but as time went on, I found myself using it less and less.' The reason he gives for stopping is

> This is going on my fourth year. At this point I must have marked close to a thousand texts [...] You're starting to see the patterns yourself [...] I still do the points where I comment on collocation and I comment on a lot of the things I did, but I no longer felt it necessary to look up 'Is this the best way to say it, is it possible to say it this way?' because I've seen it.

The repetitive nature of marking pupil assignments in Vg1 year after year has led to him being habituated, and he notes that

> I believe that if I get samfunnsfag [social studies] English or even international English [...] At that point, especially if I were to teach and grade those papers, I would definitely try my best to remember what I learned in my corpus linguistics.

For the time being, Thomas's intuition for the language is attuned to the work that the pupils are doing in a Vg1 class (which can be compared to Joy's experience: 'I know where they make their mistakes'). He does not see the need for corpora unless he starts teaching another type of English class.

5. Discussion and conclusion

The results indicate that the use of corpora may not be occurring very often in Norwegian schools. But the research question asked *How are corpora used by in-service English teachers in Norwegian schools?* and

the results show how corpora are used by some teachers. This increases our knowledge about what works for teachers in the school context. A finding from Callies's survey was that corpus-using teachers were mostly using corpora for reference (Callies 2019: 252). In the present study the same can be seen, with 12 of 33 corpus users doing this. But a difference in this study is that almost as many (11 of the 33) answered, 'I have used corpus-based materials in my teaching.' The interviews showed what this means for three of the teachers: Lars created collocation exercises with data from COCA, Thomas gave pupils the frequency of words or phrases, he created a '100 most frequent words of English' exercise, and Joy used corpora data to teach British versus American varieties. These informants exclusively use online interfaces, not offline concordancers. Interfaces that do not have 'kind of messy' (Lars) concordance lines were preferred. COCA is typically introduced in teacher education (Ebrahimi & Faghih 2016: 123; Leńko-Szymańska 2014: 278; Zareva 2017: 72), but Joy tended to use SkELL and Netspeak, and when the researcher later showed Lars and Thomas these, they were impressed by them. COCA is and has been used, but there are usability restrictions: Lars would not use it for younger pupils, and Thomas would not even use it for Vg1 (condemning it as 'web 2.0'). In terms of concordance lines, SkELL and Netspeak were preferred. The usability of online corpora interfaces like that of COCA or GloWbE, and of offline concordancers, allows for research into language, but seems to impede direct applications in the school context.

Also avoided by the informants was the building of local learner corpora. Joy did not see the need for it, and Thomas had not even thought of it until I suggested it. This was not unexpected, as Callies also did not find much compilation of learner corpora—only one informant (Callies 2019: 251, Table 5)—but the present questionnaire results showed mass collection of pupil texts, so the potential is there if the need is.

To return to the metaphor of 'bridging the gap', the lack of use of corpora has implications for teacher education, about which some remarks can be made. Language teachers need 'knowledge about the different ways in which corpora can be exploited in the classroom and the necessary skills for the application of this knowledge' (Leńko-Szymańska 2017: 217), but other factors influence teachers, such as the obstacles identified above: the differences between school levels, usability, and the lack of teacher need. These are obstacles regardless of

the corpus literacy of the teacher, even if teacher educators succeeded in making corpus linguists out of pre-service teachers. The two interviewed informants in this study who studied corpus linguistics as part of their master's degrees use corpora not habitually but only when they deem it pedagogically useful to them.

To address the 'gap' from the teaching practice 'side', rather than from the corpus linguist 'side', a starting point for corpus use among teachers may be to teach the tools and methods that seem to be already working for in-service teachers. This means further exploration of what already works, and feeding that back into teacher education. In-service teachers may find use for only a handful of activities, though, and many of the language-learning advantages of corpus use, especially in the area of pupil-corpus interaction, are lost when we look at only these informants' approaches. Yet it is a starting point, from which other suggestions for bridging the gap—corpora for curricular requirements (McEnery & Xiao 2010: 374-5) or a concordancer for the classroom (Breyer 2011: 207)—remain valid.

If more teachers used corpora even at the level the interviewed informants have, a certain degree of corpus awareness and corpus use would be established. This may be an underwhelming starting point, but 'bridging the gap' for its own sake cannot really be the ultimate aim of a corpus linguist interested in language teaching and learning. In the school context, we can begin with teachers' needs.

References

Baisa, Vít, and Vít Suchomel. 2014. SkELL—web interface for English language learning. In *Proceedings of RASLAN 2014 (recent advances in Slavonic natural language processing)*, edited by Aleš Horák and Pavel Rychlý, 63–70. Czech Republic: NLP Consulting. Details of the English Corpus for SkELL: https://www.sketchengine.eu/english-skell-corpus/

Braun, Sabine. 2007. Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL* 19(3): 307-328. doi:10.1017/S0958344007000535.

Breyer, Yvonne. 2009. Learning and teaching with corpora: reflections by student teachers. *Computer Assisted Language Learning* 22(2): 153-172. doi:10.1080/09588220902778328.

Breyer, Yvonne. 2011. *Corpora in language teaching and learning: potential, evaluation, challenges*. Berlin: Peter Lang GmbH Internationaler Verlag der Wissenschaften. Retrieved from https://www.researchgate.net/publication/264499058_Corpora_in_Language_Teaching_and_Learning_Potential_Evaluation_Challenges.

Callies, Marcus. 2019. Integrating corpus literacy into language teacher education: the case of learner corpora. In *Learner corpora and language teaching*, edited by Sandra Götz and Joybrato Mukherjee, 245-263. Amsterdam: John Benjamins Publishing Company.

Cardona, Margrete Dyvik, Anders Alvsåker Didriksen, and Anje Müller Gjesdal. 2014. Korpusbasert undervisning i fremmedspråkene: la elevens nysgjerrighet sette dagsorden. *Acta Didactica Norge* 8(2): 1-26.

Cohen, Louis, Lawrence Manion, and Keith Morrison. 2011. *Research methods in education* (7th edition). London and New York: Routledge.

Crosthwaite, Peter (ed.). 2020. Data-driven learning for the next generation: corpora and DDL for pre-tertiary learners. London: Routledge.

Davies, Mark. 2008-. The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. Available online at https://www.english-corpora.org/coca/.

Davies, Mark. 2013. Corpus of Global Web-Based English: 1.9 billion words from speakers in 20 countries (GloWbE). Available online at https://www.english-corpora.org/glowbe/.

Ebrahimi, Alice, and Esmail Faghih. 2016. Integrating corpus linguistics into online language teacher education programs. *ReCALL* 29(1): 120–135. doi:10.1017/S0958344016000070.

Education First. 2019. English proficiency index 2019. Retrieved from https://www.ef.com/__/~/media/centralefcom/epi/downloads/full-reports/v9/ef-epi-2019-english.pdf

Farr, Fiona. 2008. Evaluating the use of corpus-based instruction in a language teacher education context: perspectives from the users. *Language Awareness* 17(1): 25-43. doi:10.1080/09571736.2010.520728.

Heather, Julian, and Marie Helt. 2012. Evaluating corpus literacy training for pre-service language teachers: six case studies. *Journal of Technology and Teacher Education* 20(4): 415-440.

Koellner, Karen, and Deborah Greenblatt. 2018. 'Inservice Teacher Education' entry in Oxford Bibliographies. Retrieved from https://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0196.xml

Leńko-Szymańska, Agnieszka. 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. doi:10.1017/S095834401400010X.

Leńko-Szymańska, Agnieszka. 2017. Training teachers in data-driven learning: tackling the challenge. *Language Learning and Technology* 21(3): 217-241.

Lin, Ming Huei and Jia-Ying Lee. 2015. Data-driven learning: changing the teaching of grammar in EFL Classes. *ELT Journal* 69(3): 264-274. doi:10.1093/elt/ccv010.

McEnery, Tony, and Richard Xiao. 2010. What corpora can offer in language teaching and learning. In *Handbook of research in second language teaching and learning* (volume 2), edited by Eli Hinkel, 364-380. London and New York: Routledge.

Millar, Neil, and Ben Lehtinen. 2008. DIY learner corpora: bridging gaps between theory and practice. *The JALT CALL Journal* 4(2): 61-72.

Mukherjee, Joybrato. 2004. Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany. In *Applied corpus linguistics: a multidimensional Perspective*, edited by Ulla Connor and Thomas A. Upton, 239-250. Amsterdam and New York: Rodopi.

Røkenes, Fredrik Mørk, and Rune Johan Krumsvik. 2016. Prepared toteach ESL with ICT? A study of digital competence in Norwegian teacher education. *Computers & Education* 97: 1-20.

Römer, Ute. 2011. Corpus research applications in second language teaching. *Annual Review of Applied Linguistics* 31: 205-225. doi:10.1017/S0267190511000055.

Ståhl, Tore. 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. Retrieved from https://data.udir.no/kl06/ENG1-03.pdf?lang=eng.

Utdanningsdirektoret. 2019. English subject curriculum. Retrieved from https://www.udir.no/lk20/eng01-04

Zareva, Alla. 2017. Incorporating corpus literacy skills into TESOL teacher training. *ELT Journal* 71(1): 69-79. doi:10.1093/elt/ccw045.

Appendix 1. Questionnaire for English Teachers (English-language version)

The number of respondents for each question and option is included. The content of what respondents wrote in text fields is not included, although in the cases of Q's 11, 14, 15 and 16, it is summarized.

Q0
Dear English teacher / Kjære engelsklærer
Which language would you like to answer the questionnaire in? / På hvilket språk vil du besvare spørreskjemaet?
o English
o Norsk

Respondents: 113 in English; 97 in Norwegian

Q1 In what grade(s) do you teach English? Click on all that apply.
▢ 1
▢ 2
▢ 3
▢ 4
▢ 5
▢ 6
▢ 7
▢ 8
▢ 9
▢ 10
▢ Vg1
▢ Vg2
▢ Vg3

Respondents: 209
Grades 1-7: 58
Grades 8-10: 62
Mix of the above categories (3-10): 5
Vg: 80
Mix of Vg and grades down to 7: 4

Q2 In which county (*fylke*) do you teach?
▼ Østfold [...] Finnmark

Respondents: 206

Akershus: 44	Rogaland: 10 Nord-Trøndelag: 5
Oslo: 24	Sør-Trøndelag: 10Aust-Agder: 4
Hordaland: 19	Sogn & Fjordane: 8Troms: 4
Hedmark: 15	Vest-Agder: 8 Vestfold: 4
Buskerud: 12	Møre & Romsdal: 7Finnmark: 0
Østfold: 11	Telemark: 6
Oppland: 10	Nordland: 5

Q3 How many years have you taught in the Norwegian school system?

Respondents: 210

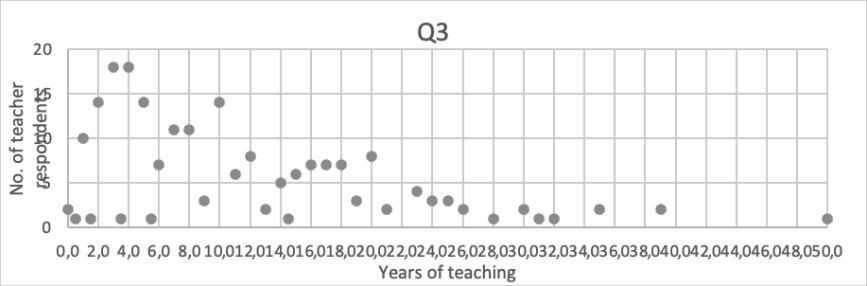


Figure 2. Scatter plot: the range is 0 (i.e. less than a year) to 50 years.

Q4 What is your approximate age? Click on one of the answers below.
o 20-29
o 30-39
o 40-49
o 50-59
o 60+

Respondents: 209
20-29: 45 30-39: 68 40-49: 57 50-59: 33 60+: 6

Q5 How do you teach English grammar? Choose any that apply to you:

☐ I teach grammar deductively (giving rules, followed by examples).

☐ I teach grammar inductively (giving examples, from which rules can be noticed or discovered).

☐ I teach grammar communicatively (I do not teach it explicitly)

☐ In another way (please specify):

Respondents: 203
Communicatively only: 36
Deductively & inductively: 32
Deductively, inductively & Communicatively: 31
Deductively only: 31
Inductively & communicatively: 30
Inductively only: 19
Deductively & communicatively: 12
Another way only: 5
Communicatively & another way: 4
Inductively & another way: 1
Deductively, communicatively & another way: 1
All four options: 1
Total selections of communicatively: 115
Total selections of inductively: 114
Total selections of deductively: 108
Total selections of another way: 12

Q6 When you teach the understanding and use of English in different situations, what kind of material do you base your teaching on? Choose all that apply to you:

- ☐ Textbooks
- ☐ Dictionaries
- ☐ English usage books
- ☐ Other English-language books (including literature)
- ☐ Newspapers/magazines
- ☐ Comics
- ☐ Online written material
- ☐ Online quizzes
- ☐ Film / TV / YouTube
- ☐ English-language song lyrics
- ☐ Radio / podcasts / audiobooks
- ☐ English speakers in the school or neighbourhood
- ☐ Field studies
- ☐ Other (please specify):

Respondents: 198
Textbooks: 180
Dictionaries: 77
English usage books: 38
Other English language books (including literature): 146
Newspapers/magazines: 105
Comics: 76
Online written material: 164
Online quizzes: 79
Film/TV/YouTube: 180
English-language song lyrics: 150
Radio/podcasts/audiobooks: 58
English speakers in the school or neighbourhood: 30
Field studies: 8
Other: 13

Q7 Do you ever base your teaching on common mistakes that your pupils make?
o Yes
o Sometimes
o No

Respondents: 196
Yes: 100
Sometimes: 88
No: 8

Display This Question: If Q7 = Yes, or Q7 = Sometimes
Q8 When you teach based on common pupil mistakes, how do you get an overview of the mistakes? Click on anything that applies to you.
▢ Personal judgement / intuition / introspection
▢ A collection of pupil texts
▢ Other (please specify):

Respondents: 188 (i.e. all who were asked)
Collection of pupil texts only: 85
Personal […] & collection of pupil texts: 60
Personal […] only: 29
Collection of pupil texts & other: 7
All three options: 5
Personal […] & other: 1
Other only: 1
Total for collection of pupil texts: 157
Total for personal […]: 95
Total for other: 14

Q9 Have you ever heard the linguistic term *corpus*? Choose from one of the answers below.

o I have never heard the linguistic term *corpus* before.
o I have heard the linguistic term *corpus* before, but I have little or no idea what it is.
o I am fairly familiar with corpus linguistics, but I have never done any practical work with corpora.
o I have already done some work with corpora.
o If you have an answer that you are sure is different to the ones above, please give it here:

Respondents: 193
Never heard the term before: 29
Have heard, but have little or no idea what it is: 75 (these respondents were asked Q14 later)
Fairly familiar with corpus linguistics, but have never done any practical work with corpora: 55 (these respondents were asked Q15 later)
Have already done some work with corpora: 34 (these respondents were asked Q's 16 & 17 later)
Different answer text field: 0

Q10 Do you know what any of these terms mean?

	Yes	Not sure	No
No collocation	o	o	o
Colligation	o	o	o
Concordance	o	o	o
Concordancer	o	o	o
Frequency list	o	o	o
Key word in context (KWIC)	o	o	o
Part of speech (POS)	o	o	o

Respondents: 191

No collocation	Yes: 107	Not sure: 39	No: 45	
Colligation	Yes: 17	Not sure: 72	No: 102	
Concordance	Yes: 101	Not sure: 40	No: 50	
Concordancer	Yes: 34	Not sure: 63	No: 92	No answer: 2
Frequency list	Yes: 151	Not sure: 21	No: 19	
KWIC	Yes: 112	Not sure: 42	No: 37	
POS	Yes: 157	Not sure: 21	No: 13	

Display This Question: If any part of Q10 was answered
Q11 Where did you encounter the term(s)?

Respondents: 125 (of 186 asked)
These are the number of mentions for each main source.
Studies: 99
Books/texts: 12
Work/colleagues: 8
Internet: 6

Q12 Do you know any of these web tools? Select any you know. (The screenshots may take time to load.)

- ☐ SkELL
- ☐ Using English
- ☐ BYU corpora
- ☐ BNC Simple Search

Respondents: 64
SkELL identified by: 8 Using English identified by: 40
BYU corpora identified by: 24 BNC Simple Search identified by: 24

The 9 non-interviewed corpus users.
SKELL identified by: 1 Using English identified by: 6
BYU corpora identified by: 3 BNC Simple Search identified by: 5

Q13 Do you know other web tools similar to the ones above? If so, please name them.

Respondents: 15

Display This Question: If Q9 = I have heard the linguistic term corpus before, but I have little or no idea what it is.
Q14 Above, you answered that you have heard the linguistic term *corpus* before, but you have little or no idea what it is. Where have you encountered the term?

Respondents: 51 (of 75 asked)
These are the number of mentions for each main source.
Studies: 20 Internet: 11Books: 5

Display This Question: If Q9 = I am fairly familiar with corpus linguistics, but I have never done any practical work with corpora.

Q15 You are fairly familiar with corpus linguistics, but you have never done any practical work with corpora. How did you learn about corpus linguistics?

Respondents: 48 (of 55 asked)
These are the number of mentions for each main source.
Studies: 38
Indirectly/conversation: 3
Reading/articles: 3

Display This Question: If Q9 = I have already done some work with corpora.
Q16 You have already done some work with corpora. In what context did you do this? Please click on any context below that applies to you.

- ▢ I worked with corpora as part of teacher training.
- ▢ I worked with corpora in a course (not a teacher training course).
- ▢ I have used corpora to check acceptability of use when in doubt (or when marking).
- ▢ I have used corpora-based materials in my teaching.
- ▢ I have introduced corpora to pupils.
- ▢ Other (please specify):

Respondents: 33 (of 34 asked)
As part of teacher training: 25
In a course (not teacher training): 9
To check acceptability of use when in doubt (or when marking): 12
Used corpora-based materials in teaching: 11
Introduced corpora to pupils: 6
Other: 2 (master's courses)

Display This Question: If Q9 = I have already done some work with corpora.
Q17 What have you used corpora for? Please click on any use below that applies to you.

- ☐ Grammar
- ☐ Vocabulary
- ☐ Spelling
- ☐ Idiom
- ☐ Authentic dialogue
- ☐ Learner language
- ☐ Other (please specify):

Respondents: 31 (of 34 asked)
Grammar: 15
Vocabulary: 22
Spelling: 2
Idiom: 19
Authentic dialogue: 15
Learner language: 8
Other: 3

Display This Question: If Q16 = I have used corpora-based materials in my teaching, or Q16 = I have introduced corpora to pupils, or Q16 = Other (please specify): Is Not Empty
Q18 Your experience is of interest to our researcher! Would you be willing to be interviewed about your work with corpora?

- o Yes. Here is an email address:
- o No.

Respondents: 14 (i.e. all who were asked)
Yes: 4
No: 10

Appendix 2. Interview guides.

Interview guide – informant 'Joy', 2018.

Focus: Corpus linguistics

Research question: How familiar are teachers of English in Norway with corpus linguistics?

Question:	Answer:
Why did you become an English teacher?	
How have you worked with corpora in the context of checking English usage?	
How have you worked with corpora in the context of using corpus-based materials in teaching?	
How have you introduced corpora to pupils?	
How have you used corpora for vocabulary?	
How have you used corpora for idiom?	
How have you used corpora for learner language?	
How have you used corpora for grammar?	
How have you used corpora for spelling?	
How have you used corpora for varieties of English?	
When you answered the question 'When you teach the understanding and use of English in different situations, what kind of material do you base you teaching on', you didn't mention corpora. Have you any idea why it didn't come to mind?	

Interview guide – informant 'Lars', 2018.

Focus: Corpus linguistics

Research question: How familiar are teachers of English in Norway with corpus linguistics?

Question:	Answer:
Why did you become an English teacher?	
What was your experience of encountering corpus linguistics in your master's degree?	
How did you work with corpora in the context of teacher training?	
How have you worked with corpora in the context of checking English usage?	
How have you worked with corpora in the context of using corpus-based materials in teaching?	
How have you introduced corpora to pupils?	
How have you used corpora for vocabulary?	
How have you used corpora for idiom?	
How have you used corpora for authentic dialogue?	
How have you used corpora for grammar?	
When you answered the question 'When you teach the understanding and use of English in different situations, what kind of material do you base you teaching on', you didn't mention corpora. Have you any idea why it didn't come to mind?	

Interview guide – informant 'Thomas', 2018.

Focus: Corpus linguistics

Research question: How familiar are teachers of English in Norway with corpus linguistics?

Question:	Answer:
Why did you become an English teacher?	
What was your experience of encountering corpus linguistics in your master's degree?	
How did you work with corpora in the context of teacher training?	
How have you worked with corpora in the context of using corpus-based materials in teaching?	
How have you used corpora for vocabulary?	
How have you used corpora for idiom?	
How have you used corpora for authentic dialogue?	
When you answered the question 'When you teach the understanding and use of English in different situations, what kind of material do you base you teaching on', you didn't mention corpora. Have you any idea why it didn't come to mind?	
How do you collect pupil texts?	

2

Nordic Journal of Language Teaching and Learning

NJLTL Vol. 9 No. 2 (2021) 10.46364/njltl.v9i2.933

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 applied[1] 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-service[2] 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 & Faghih, 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 & Faghih, 2016, p. 128).[3]

Usability. Pre-service teachers perceived corpus software as difficult to use (Ebrahimi & Faghih, 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 & Faghih, 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 & Faghih, 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 & Faghih, 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 & Faghih, 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 LancsLex, 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 pretedermined; 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 & Faghih, 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-corpus 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 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/kl06/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 ompetence 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

3

Corpus Exercises for Lower Secondary English

1. Introduction

This article shows how suggested corpus exercises are warranted by the current English subject curriculum for lower secondary school (years 8-10[1]) in Norway. There has been research on the benefits of corpora for language learning, for example a meta-analysis by Boulton and Cobb (2017), and there exists the potential of using language corpora in schools (e.g. Braun 2007; Crosthwaite 2020), yet corpora do not seem to be used very often by teachers (Callies 2019; Kavanagh 2021a). Internationally, researchers have found there to be a 'gap' between corpus linguistics and teaching practice, and have discussed the difficulty of 'bridging' the gap (Mukherjee 2004; Breyer 2011: 146). One attempt to further the use of corpora has been the goal of educating teachers and teacher-education students to become corpus linguists themselves. Various studies have evaluated corpus instruction in teacher education (Breyer 2009; Ebrahimi & Faghih 2017; Farr 2008; Leńko-Szymańska 2014, 2017; Zareva 2017), but a particular insight from that material is that a single-semester course in corpus instruction is not enough to make teachers confident users of corpora in teaching (Leńko-Szymańska 2014: 272). This seems to be the case in Norway (Karlsen & Monsen 2020: 132). Even when teachers have had more than a semester's experience with corpora in teacher education, this still does not seem to lead to the use of corpora in their teaching (Kavanagh 2021a: 10, 12, 18). Meunier points out that applied linguists have not tended to explicitly verbalize how corpus activities 'meet curricular demands or expectations', and that is 'one of the reasons for… lack of uptake in teaching contexts other than university-level courses' - the activities need to 'fit in the broader curricular context in which they are carried out' (Meunier 2022: 348). This article suggests corpus exercises that teachers of English in Norway may find accessible, and that relate to the curriculum. The aim of this article is to answer the research question *How can corpora be used in the English subject in lower secondary school with the current curriculum in Norway?*

There are five suggested corpus exercises for vocabulary in this article. This is not an enormous number of exercises, but teacher educators may find this study useful: the exercises can serve as examples of how teachers can use corpora in the English subject, and their student teachers do not have to have experience of corpus linguistics in order to try out the exercises. Any in-service lower secondary English teachers reading this may like to experiment with the exercises.[2]

[1] Normally, from the year a pupil turns 13 to the year a pupil turns 16.

[2] In-service 'designates 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).

The remainder of this introduction clarifies what kind of corpus use is involved in this article's suggested corpus exercises (section 1.1), why the exercises are linked to pre-existing textbook exercises (section 1.2), why the focus is on lower secondary school (section 1.3), and what the obstacles that must be overcome are (section 1.4).

1.1. The corpus use involved in this study

The corpus use envisaged here is direct pedagogical corpus applications, which can be either 'teacher-corpus interaction', 'learner-corpus interaction' or both; this differs from 'indirect applications', which affect the content of reference books, textbooks, and syllabi (Römer 2011: 207). Römer describes 'direct applications' as 'hands on for learners and teachers (data-driven learning)' (ibid.). This article's suggested corpus exercises have been designed for learner-corpus interaction. They involve 'hands-on' computer use by pupils.[3]

The suggested exercises in this article can be considered data-driven learning (DDL) in the broad sense of corpora used 'for pedagogical purposes' (Gilquin & Granger 2022: 430), but may not be considered DDL when the term is used to mean inductive learning only. The exercises have both deductive and inductive aspects. Pupils need to have language concepts (synonyms, collocations – see section 4.2) explained first, which is deductive learning, but there are instances when pupils choose which words to investigate, or when pupils are expected to discover semantic patterns for themselves, and these aspects of the exercises are inductive learning.

Guidance from the teacher need not be absent in hands-on interaction: 'the teacher may decide the points to cover, devising step-by-step tasks with set queries leading to pre-established outcomes, and monitoring performance at all stages' (Boulton 2012: 154). The role of the teacher in DDL was initially conceived of as a provider of 'a context in which the learner can develop strategies for discovery' (Johns 1991: 1), which may seem vague. Breyer details what is actually involved for the teacher in teaching and learning with corpora: the teacher needs to, first, 'possess a degree of "corpus literacy"[4] in order to teach with these materials and integrate them meaningfully into the classroom. The teacher has to guide the learners who will most likely be novice users of corpora'; second, 'assess the materials for their appropriateness in the respective learning context; that is language proficiency level, suitable vocabulary, integration into the curriculum and so forth'; and potentially third, teachers need to 'create their own materials …

[3] 'Hands-off' corpus interaction, in which pupils use prepared materials from a corpus instead of interacting with a corpus using a computer, is also possible, has also be referred to as 'data-driven learning' (DDL) (Vyatkina 2016), and has its advantages (Boulton 2012: 154-155), including a greater degree of guidance from the teacher.

[4] For a detailed definition of corpus literacy, see Callies 2019: 247.

finding or creating an appropriate corpus, acquiring and learning how to work with concordancing software, creating meaningful exercises, and producing worksheets' (Breyer 2009: 156). There is a guiding role for the teacher in this article's hands-on exercises. To use the exercises with pupils, the teacher does not need to be a corpus linguist, and how much of the above the teacher needs to do will be explained in section 5.3.

1.2. The exercises' link to pre-existing textbook exercises

The suggested corpus exercises in this article adapt exercises from lower-secondary-school textbooks. This is because corpora as resources would be especially useful if they could enhance, or were more useful than, the resources that are already available for teachers. Recently published textbooks are such resources, linked to the curriculum that came into effect in August 2020 (Norwegian Directorate for Education and Training 2019).[5] There is a strong textbook focus in primary and lower secondary school levels in Norway. At years 5-10, 70% of surveyed English teachers said they primarily use paper-based textbooks as resources (Gilje et al. 2016: 52). The suggested corpus exercises can replace, enhance or revise textbook exercises. That the exercises correspond to textbook exercises shows that corpora can link to the curriculum in a similar way to the textbooks, or even more so (see section 3). Additionally, for choosing which language topics to cover in the suggested corpus exercises, an investigation of textbooks helps identify prominent topics (see section 4.2).

1.3. The focus on lower secondary school (years 8-10)

Given the curricular competence aims after year 10 (Norwegian Directorate for Education and Training 2019: 8-9), and the views of lower secondary teachers in previous research (Kavanagh 2021a; Kavanagh 2021b), lower secondary school seemed the best choice of focus for this research, because lower secondary competence aims, and lower secondary teachers, seemed to have comparatively more to say about language work in the English subject. However, were there space for a wider focus, primary and upper secondary school levels would also have been included in this study.

1.4. Obstacles: the reasons for lack of direct applications

Teacher informants themselves have suggested various reasons why direct applications are not more commonly used. First, corpus technology can lack usability ('user friendliness') from the teacher perspective (Karlsen & Monsen 2020: 131; Kavanagh 2021a: 17-18). Second, teachers

[5] There also exist digital resources for teachers in Norway, but the present author was not given access to them.

may perceive corpora as tools for linguists, unsuitable or too expensive or inaccessible for schools (Karlsen & Monsen 2020: 134). Third, teachers may lack digital skills (Kavanagh 2021b: 99), or the pupils may do so (Karlsen & Monsen 2020: 135), the idea that the current generation in education would be 'digital natives' seeming to apply in reality only to a minority of pupils (Ståhl 2017). Fourth, teachers may not be inclined to use linguistic materials, either in the lower years if they perceive pupils' language mistakes to be 'obvious' (Kavanagh 2021a: 18; Kavanagh 2021b: 100), or in the higher years because teachers may become more topic-focused than language-focused (Karlsen & Monsen 2020: 138, 141, 142, 144; Kavanagh 2021a: 17). Fifth, while classroom work with corpora can be either deductive or inductive (Breyer 2011: 52-54; Liu & Lei 2017: 31-34; Pinto et al. 2023), in both cases this means explicit language learning, which may clash with teacher beliefs about implicit language learning (Karlsen & Monsen 2020: 139). The effects of these five obstacles need to be mitigated if teachers are to use corpora. In the design of the suggested corpus exercises, it is hoped that the choice of corpora (see section 4) can mitigate some of the effects of the first three obstacles, while the curriculum itself can mitigate the effects of the last two (see section 3).

The rest of this article is structured as follows. Section 2 discusses previous research relevant to the suggested corpus exercises, section 3 shows how a corpus-based approach can help achieve learning objectives in the curriculum, section 4 explains the choices of corpora and language topics, section 5 discusses the corpus exercises, and section 6 contains concluding remarks.

2. Previous research relevant to the suggested corpus exercises

This section discusses previous research on what teachers have used corpora for, and which corpora teachers have used. The reason to explore this research when designing corpus exercises is to build upon what some teachers already do, so that corpus exercises are designed that teachers will more likely find useful.

2.1. Previous research on what teachers have used corpora for

Farr (2008), Zareva (2017), Callies (2019) and Kavanagh (2021a) have investigated what some teachers knowledgeable of corpus linguistics use corpora for. Farr's informants' list of what uses they 'might make of corpora' is very long and will not be reproduced here (Farr 2008: 38); but the list of 'areas of teaching' indicated by Zareva's informants for the use of corpora is short enough: 'grammar, academic vocabulary, and the use of words and collocates' (Zareva 2017: 75). Both Farr's and Zareva's studies had a mix of in-service and pre-service teachers, so they could not be certain what their informants actually used in teaching (Farr 2008: 39; Zareva 2017: 75). Callies's

and Kavanagh's informants were exclusively in-service teachers, so these studies do not share the same uncertainty. Most of Callies's informants who used corpora said they did so for reference, 'for example when marking or checking language acceptability, while corpora are hardly, if at all, used in learner-centred activities' (Callies 2019: 252). This use of corpora for reference was mirrored in Kavanagh's informants in Norway (Kavanagh 2021a: 19), but some of the informants (n=12) said they used corpora in teaching,[6] for: vocabulary (n=11), idiom (n=10), authentic dialogue (n=8), grammar (n=6), learner language (n=5), spelling (n=2), and other uses by single informants (Kavanagh 2021a: 12). In follow-up interviews with three of these informants, it was established which uses of corpora were teacher-corpus interaction, which were learner-corpus interaction, and which particular corpora were used for each purpose (Kavanagh 2021a: 13-15, Tables 1 and 2). That so many informants see the usefulness of corpora for vocabulary teaching and learning is encouraging. Hands-on work with corpora has led to measurable vocabulary learning (Cobb 1997; Lee, Warschauer & Lee 2019). However, in a study with pre-service teacher informants, concern was expressed about informants 'addressing mainly vocabulary and collocations and not any other language features', noting that 'it was impossible to establish if this focus … was a result of their lack of confidence in analysing corpus data for other features' (Leńko-Szymańska 2017: 233). This could be the case with vocabulary teaching and learning in Kavanagh's and Zareva's studies. Taking all of this into consideration, section 4.2 will explain the choice of what corpora are used for in this study's suggested exercises.

2.2. Corpora that teachers have used: previous research

'Pedagogically motivated' corpora (McEnery & Xiao 2010: 374-375) have been suggested, but successfully relevant corpora for Norwegian schools do not yet seem to exist. A study of the use, in two upper secondary school classrooms, of a set of corpora (BACKBONE), designed and assembled for pedagogic purposes, concluded that the corpora struggled to be interesting and relevant to pupils (Farr & Karlsen 2022; Karlsen in preparation). If there are no suitable pedagogically motivated corpora, the use of general corpora in teaching must be considered. Of the studies mentioned in section 2.1., three have data about which corpora the informants used. Farr's and Zareva's informants were enrolled in educational programmes, and it can be assumed they used the corpora introduced to them in that context. Farr's informants had access to: the British National Corpus (BNC), the American National Corpus (ANC), the Corpus of Spoken Professional English, the ICAME collections, the Limerick Corpus of Irish English (L-CIE), and

[6] 'In teaching' in this context means informants (n=12) who answered 'I have used corpus-based materials in my teaching' and/or 'I have introduced corpora to pupils' in a questionnaire (n=210) (Kavanagh 2021a: 10-11).

'available online collections or samplers, e.g. the Cobuild sampler' (Farr 2008: 31). Farr's informants also had access to offline corpus software Wordsmith Tools, Monoconc and Paraconc (ibid.). Zareva's informants were introduced to two corpora: the Corpus of Contemporary American English (COCA) (Davies 2008-), and the Michigan Corpus of Academic Spoken English (MICASE). Kavanagh's corpus-using informants did not use offline corpus software; they all used online corpus interfaces (Kavanagh 2021a: 12). These were: SKELL, a billion-word text corpus that consists of 'sentences sorted according to their text quality', from selected corpora and websites (Baisa & Suchomel 2014); Netspeak (https://netspeak.org/), which searches Google Books; and English-Corpora.org (Davies 2002-), from where COCA, and Global Web-Based English (GloWbE) (Davies 2013), can be accessed (Kavanagh 2021a: 13-15). Two informants who used COCA did not think highly of its usability, and would not use it with most pupils (ibid.: 19). Taking all of this into consideration, section 4.1 will explain the choice of corpora used in this study's suggested exercises.

3. The 2020 curriculum

This section shows how a corpus-based approach can help achieve important learning objectives in both Norway's core curriculum and its English subject curriculum. The *Core Curriculum – values and principles for primary and secondary education* (Norwegian Directorate for Education and Training 2017) provides the values and principles of education, while a subject curriculum describes the content and goals of the subject.

3.1. The core curriculum

Corpus use in English language teaching is warranted by important concepts in the *Core Curriculum*, namely critical thinking and in-depth learning. Critical thinking 'means applying reason in an inquisitive and systematic way when working with specific practical challenges, phenomena, expressions and forms of knowledge' (Norwegian Directorate for Education and Training: 6), and includes thinking critically about sources, with 'room for uncertainty and unpredictability' (ibid.). This can apply to linguistic data in a language subject. Corpus searches present practical challenges, and there is uncertainty and unpredictability in language data. A practical challenge may be, for example, how to formulate a search term for retrieving specific data from a corpus search; and uncertainty and unpredictability can occur when corpus data show tendencies in the usage of a word or phrase, variation in usage, and the frequency of usage – the data will not give a prescription for usage. Drawing conclusions from such data requires critical thinking.

In-depth learning[7] can be 'defined as students' understanding of concepts and the relations between them, students relating new ideas to familiar concepts and principles in order for the new understanding to be used in problem-solving in new and unfamiliar situations' (Burner 2020: 55). In practice, in-depth learning can mean many things, one of which is that pupils can make the choice to go in-depth in a subject, with the school providing sufficient time for specialization, as well as support and guidance (Norges Offentlige Utredninger 2015: 11; Norwegian Directorate for Education and Training 2017: 7, 12, 17). In-depth learning can apply to a language subject, or to cross-curricular work that includes a language subject. Introducing pupils to corpora gives them the opportunity to begin an endless investigation into words and phrases. Once learners understand the search field of a corpus interface, they can type anything into it, which shows the possibility of open-ended corpus research.

3.2. The English subject curriculum

Corpora can help fulfil curricular requirements for a focus on language in general, and explicit language learning in particular. As mentioned above as an obstacle to corpus use (section 1.4), teachers in the higher years of the English subject in schools in Norway may have been more topic-focused (e.g. on sociocultural topics) than language-focused. This may have been due to the previous curriculum (Norwegian Directorate for Education and Training 2013), whereas according to the current curriculum, language topics of the subject should be taught at every level (Norwegian Directorate for Education and Training 2019: 5-12), which makes the use of the linguistic data that corpora can provide very relevant. The English subject curriculum refers, in its 'core elements' section, to 'knowledge of English as a system' as part of language learning, and 'vocabulary, word structure, [and] syntax' are among what must be learnt (ibid.: 2). This is reinforced in the 'competence aims after year 10' specific to lower secondary (ibid.: 8-9). It is difficult not to see the reference to language as a system as an expectation that some language teaching will be explicit, in which case this curriculum seems open to the use of linguistic data that corpora can provide, and teachers can be open to using such material. Lower-secondary textbooks have many explicit language exercises, for example, *Enter 10 Engelsk* (i.e. for year 10) has a task in which the pupils must find phrasal verbs (Diskin & Winsvold 2021: 23). The linguistic data of a corpus is another way to introduce pupils to language elements explicitly. For example, with a selection of phrasal verbs, corpus data could show which are commonly used.

It was also mentioned above, as an obstacle to corpus use (section 1.4), that teachers may have beliefs about implicit language learning, whereas the subject curriculum appears to imply that

[7] Sometimes translated from Norwegian *dybdelæring* as 'deep learning'.

some explicit learning is necessary. In practice, for most teachers there may be no clash in facilitating both implicit and explicit learning, because for decades in Norway a version of communicative language teaching has been practised in which learning a language has not been seen as implicit as learning one's first language, and 'facilitating the use of English for communicative purposes' has not been seen 'as the only route to learning the language', thus 'an eclectic approach' has been 'encouraged' (Skulstad 2020: 56). A 'focus on form' is acceptable within a communicative approach, as long as form 'is studied in context' (Leńko-Szymańska & Boulton 2015: 5).

There is also a digital aspect to the English subject curriculum. The use of corpora in teaching involves digital skills, which is one of the basic language skills identified in the English subject curriculum. In the competence aims for year 10, 'The pupil is expected to be able to… use different digital resources and other aids in language learning' (ibid.: 8). The use of corpora, which can be considered 'digital resources' for learning, can help fulfil this.

4. The choices of corpora and language topics

This section discusses the choice of corpora for the suggested corpus exercises, and the choice of language topics in the exercises. These choices were made in light of what previous research seems to indicate that teachers find useful (section 2). The rationale behind this is to build on what some teachers already work with.

4.1. The choice of corpora for the suggested corpus exercises

The corpora chosen for the suggested corpus exercises were SKELL and Netspeak, for the following reasons. There are indications that teachers may be willing to use them in the English subject, details of which are given in the paragraphs below. Teachers may perceive these as more easily usable than other corpora, they may require fewer prior digital skills from teachers and pupils, and they may not give the impression of being tools for linguists. Both SKELL and Netspeak are free to use, and do not require users to create an account. Taking all of this together, SKELL and Netspeak seem like they can mitigate the first three obstacles to corpus use mentioned earlier (section 1.4).

An English teacher in Norway was interviewed about her use of SKELL, and her introduction of it to pupils in upper secondary school: in both teacher-corpus and learner-corpus interaction, it was used for the purpose of vocabulary teaching and learning (Kavanagh 2021a: 13-16). Other English teachers in Norway interviewed about SKELL had some positive impressions on being introduced to it (Kavanagh 2021a: 19; Kavanagh 2021b: 97-99). For one teacher, it seemed easy

to use because it merely involved opening a web page and typing; another teacher talked of it as an alternative to dictionaries; and a third said SKELL search results, which are presented as full sentences, were more readable than the typical concordance lines of a corpus, therefore easier to use in teaching (Kavanagh 2021b: 97-99).[8]

The same teacher who used SKELL in upper secondary used Netspeak, and introduced it to pupils: in teacher-corpus interaction, the purpose was vocabulary teaching and learning, and in learner-corpus interaction, it was for vocabulary and grammar (Kavanagh 2021a: 13-16). Other English teachers in Norway introduced to Netspeak had some positive impressions (Kavanagh 2021a: 19; Kavanagh 2021b: 97-99). For one teacher, it was easy to use in the same manner as SKELL; and another teacher said the 'finds one word' feature of Netspeak could be usable for pupils as low as year 6 (Kavanagh 2021b: 97-99).

However, teachers introduced to SKELL and Netspeak have noted their limitations. A teacher thought SKELL's 'Examples' (its most basic feature) would be difficult for pupils who, at year 8, had yet to develop dictionary skills,[9] while both she and an upper secondary teacher thought its 'Word sketch' feature would not be understandable to pupils (Kavanagh 2021b: 99-100). The same upper secondary teacher also criticized the context-free display of SKELL's example sentences, and the order in which they appear (Kavanagh 2021b: 100). A teacher who suggested year 8 pupils could use Netspeak's 'check the order' feature also stated that this would not be for all pupils, but only for 'curious language learners' (Kavanagh 2021b: 97, 100). These opinions were taken into consideration when designing this article's suggested corpus exercises, which are for years 8-10; for example, the use of SKELL's 'Word sketch' and Netspeak's 'check the order' are avoided.

4.2. The choice of language topics in the suggested corpus exercises: synonyms and collocations

It was decided that the suggested corpus exercises focus on two language elements: synonyms and collocations. This is for the following reasons: both synonyms and collocations are vocabulary topics (section 4.2.1), both topics relate to the curriculum (section 4.2.2), and exercises for both appear in lower secondary school textbooks (section 4.2.3).

[8] There has been an experiment where students scored higher in a test on concordance questions than on full sentence questions (Boulton 2009: 47), but the students were at an engineering college and had the average age of 18.5 (ibid.: 41).

[9] This teacher was interviewed when the previous English subject curriculum was in effect (Kavanagh 2021b: 95). The current curriculum specifies, with competence aims for the end of years 4 and 7, that pupils should have dictionary skills earlier than this (Norwegian Directorate for Education and Training 2019: 6, 7).

4.2.1. Synonyms and collocations are vocabulary topics

Previous studies show teacher awareness that corpora can be used for vocabulary teaching and learning (Kavanagh 2021a; Zareva 2017). Synonyms and collocations are both vocabulary topics. It was mentioned above that teachers might choose to address vocabulary language elements because of inexperience with corpora (Leńko-Szymańska 2017: 233), but this article's suggested corpus exercises are intended to be as useful to teachers who have no experience with corpora as they are to those who do have experience. Vocabulary teaching and learning can be a first step in pedagogical corpus use, and exercises can follow for other uses. Furthermore, the term 'vocabulary' is used in the curricular competence aims only for lower secondary, after year 7 and after year 10 (Norwegian Directorate for Education and Training 2019: 7-9).

Synonyms is a topic that has been suggested as 'difficult… for learners' and a potentially useful area for learning with corpora (Breyer 2011: 56). Learning collocations involves identifying collocates, and it is difficult for learners to rely on their intuition: '[m]ost linguists today agree that the only way to reliably identify the collocates of a given word or phrase is to study patterns of co-occurrence in a text corpus' (McEnery & Hardie 2012: 123). This is a task for which corpora are particularly suited. SKELL and Netspeak can be used for vocabulary teaching and learning (Kavanagh 2021a: 13-16).

4.2.2. Synonyms and collocations are topics that relate to the curriculum

Specific language elements can be linked to specific year 10 competence aims; a textbook for English teacher education provides examples of these links (Flognfeldt & Lund 2021: 89-90, 203-204). The authors connect 'words and lexical chunks (including sentence structures) related to different topics and functions' (ibid.: 89) to two year 10 curricular aims: 'the pupil is expected to be able to… express himself or herself with fluency and coherence with a varied vocabulary and idiomatic expressions adapted to the purpose, receiver and situation' and 'listen to and understand words and expressions in variants of English' (Norwegian Directorate for Education and Training 2019: 8). In proposing synonyms exercises and collocations exercises, both of which can be seen as part of work with words and lexical chunks, the present study interprets work with synonyms as being linked to the curricular term 'varied vocabulary', and work with collocations as being linked to the curricular terms 'idiomatic expressions' and 'adapted to the purpose, receiver and situation'. Exercises with synonyms across English varieties, for example Irish compared with British synonyms (Pettersen & Røkaas 2020: 208), are interpreted as being linked to the curricular term 'variants of English'.

4.2.3. Synonyms and collocations exercises are in lower-secondary textbooks

In lower secondary school textbooks, synonyms appears as a language topic in all of the lower-secondary textbook series *Engelsk, Enter* and *Stages*. Collocations is a language topic in two of the series, *Enter* and *Stages*. The suggested corpus exercises correspond with types of exercises found in textbooks. They therefore link to the curriculum in the same way, with (presumably) the same teaching aims as the textbook authors.

The English textbooks for lower secondary school that were explored for this study were: *Engelsk 8 fra Cappelen Damm Student's Book* (Madsen & Mohammad-Roe 2020), *Engelsk 9 fra Cappelen Damm Student's Book* (Haegi, Madsen & Mohammad-Roe 2020), *Engelsk 10 fra Cappelen Damm Student's Book* (Haegi, Madsen & Mohammad-Roe 2021), *Enter 8 Engelsk* (Diskin, Kasbo & Winsvold 2020a), *Enter 10 Engelsk* (Diskin & Winsvold 2021),[10] *Enter 8-10 Basic Skills*[11] (Diskin, Winsvold & Kasbo 2020b), *Stages 8 Engelsk for ungdomstrinnet* (Pettersen & Røkaas 2020), *Stages 9 Engelsk for ungdomstrinnet* (Pettersen & Røkaas 2021a), and *Stages 10 Engelsk for ungdomstrinnet* (Pettersen & Røkaas 2021b).

The *Enter* series prominently introduces the concept of synonyms (Diskin & Winsvold 2021: 201; Diskin, Winsvold & Kasbo 2020b: 74; both reproduced here in Appendix 1A). Chapter 5 of *Engelsk 8 fra Cappelen Damm Student's Book*, and chapter 1 of *Engelsk 10 fra Cappelen Damm Student's Book* state that synonyms is one of the 'language and grammar' learning targets (Madsen & Mohammad-Roe 2020: 187; Haegi, Madsen & Mohammad-Roe 2021: 7).

These are the eight types of synonyms exercises found in the books:

1 Find synonyms for words (Madsen & Mohammad-Roe 2020: 216, 233; Diskin, Kasbo & Winsvold 2020a: 27, 113; Diskin & Winsvold 2021: 41, 53, 145, 169, 195, 201; Diskin, Winsvold & Kasbo 2020b: 75; Pettersen & Røkaas 2020: 32, 259; 2021a: 156, 276, 277; 2021b: 13, 272), for adjectives in particular (Haegi, Madsen & Mohammad-Roe 2020: 197; 2021: 44), or for multi-word expressions (Diskin, Kasbo & Winsvold 2020a: 153); in some of these exercises, texts are provided in which the synonyms can be found (Diskin, Kasbo & Winsvold 2020a: 27; Diskin & Winsvold 2021: 145, 195; Pettersen & Røkaas 2021a: 156; 2021b: 13), most simply with 'replace underlined words with synonyms from a list' (Pettersen & Røkaas 2021a: 78). One exercise of

[10] *Enter 9 Engelsk* (Diskin & Winsvold 2020) was also explored, but it does not contain synonyms exercises or collocations exercises.

[11] This is a 'grammar and language' book used alongside *Enter* textbooks through the years of lower secondary school.

this type (Madsen & Mohammad-Roe 2020: 233) is reproduced in Appendix 1B. Textbook level: years 8-10.

2 Define 'synonym' and collect examples (Madsen & Mohammad-Roe 2020: 232). Reproduced in the Appendix 1C. Textbook level: year 8.

3 Explore the meanings of words that can replace *say* (Pettersen & Røkaas 2021b: 28, 177). One of these exercises has two parts: in (a), the pupil relates two synonyms of *say* to a short story on the preceding pages; in (b), twenty synonyms of *say* are displayed in a 'verb wheel' and pupils are instructed to use the verbs 'in sentences' (Pettersen & Røkaas 2021b: 28); this is reproduced in Appendix 1D, without the short story). Textbook level: year 10.

4 Consider why an author did not use more common synonyms (Diskin & Winsvold 2021: 201). Textbook level: year 10.

5 Match British English words with synonyms from other national varieties (Diskin, Kasbo & Winsvold 2020a: 64; Pettersen & Røkaas 2020: 208). Textbook level: year 8.

6 Rewrite old-fashioned sentences in a modern style (Diskin, Kasbo & Winsvold 2020a: 215). Textbook level: year 8.

7 Use a Norwegian dictionary to see how many English translations of a word there are, and to consider whether they are synonyms (Diskin, Winsvold & Kasbo 2020b: 75). Textbook level: years 8-10.

8 Use synonyms as clues in a vocabulary game called 'guess my word' (Pettersen & Røkaas 2021b: 257). Textbook level: year 10.

Pupils are also directed to use a thesaurus (Diskin & Winsvold 2021: 201), an online thesaurus (Diskin, Winsvold & Kasbo 2020b: 74; Pettersen & Røkaas 2020: 32, 70; 2021a: 276, 277; 2021b: 13, 272), or a dictionary (Diskin, Winsvold & Kasbo 2020b: 74, 75; Pettersen & Røkaas 2020: 259), and to collect synonyms in a language journal (Diskin, Winsvold & Kasbo 2020b: 75).

Collocations exercises are found in *Enter 10* and *Stages 10*. Pupils are introduced to the concept in both book series (Diskin, Winsvold & Kasbo 2020b: 76; Pettersen & Røkaas 2021b: 226; both reproduced in Appendix 1E). These are the exercises:

1 'Expressions', e.g. *throw a switch* and *surge of power*, are presented to pupils, following a short story in which they appear, as examples of collocations. Pupils are to explain the meanings of the expressions, and consider why 'the use of collocations like this improves your texts' (Diskin & Winsvold 2021: 154-155). Reproduced in Appendix 1F, without the short story. Textbook level: year 10.

2 A gap-fill exercise, 'Choose the best word to complete each collocation', choosing between *strong, deep* and *heavy* (Pettersen & Røkaas 2021b: 226). Pupils are given the semantic pattern of the nouns the adjectives collocate with. See Appendix 1G. Textbook level: year 10.

Pupils are also directed to start a collection of collocations (Diskin, Winsvold & Kasbo 2020b: 77).

5. The suggested corpus exercises

This section discusses the suggested corpus exercises designed for this article. There are five exercises, three of which are for the synonyms topic (section 5.1) and two for collocations (section 5.2). Exercises 1 and 5 contain multiple tasks. The teacher's role in the exercises is also described (section 5.3).

Each exercise involves learner-corpus interaction with SKELL and Netspeak. The most basic features of SKELL and Netspeak (to which the online interfaces default) are used in the exercises, namely the 'Examples' feature of SKELL, which provides the user with example sentences containing their search term, and the 'finds one word' feature of Netspeak, in which a question mark is used to find all words that would fill a particular slot in a sentence. In addition, the 'Similar words' feature of SKELL and the 'finds similar words' feature of Netspeak are used, because they are useful for finding synonyms.

As with the textbook exercises, the suggested corpus exercises are designed to be used after it has been explained to pupils what synonyms are and what collocations are. There were eight types of synonyms exercises in the textbooks (section 4.2.3). Three of these have been adapted[12] into corpus exercises (Appendix 2, sections 2A, 2B and 2C), and section 5.1 will also comment on how corpus use can assist with the other five types. There were two collocations exercises in the textbooks, both of which have been adapted into corpus exercises (Appendix 2, sections 2D and 2E). Additionally, a combined synonyms and collocations exercise is suggested, at the end of section 5.2.

5.1. Synonyms exercises

Exercise 1. Corpus exercise for years 8-10 (Appendix 2A).

[12] 'Adapted': for each exercise discussed, in sections 5.1 and 5.2, it will be indicated whether it is designed to replace, enhance or revise one of the textbook exercises described earlier in section 4.2.3.

This corpus exercise replaces the type of textbook synonyms exercise for years 8-10, the most common synonyms exercise in the textbooks: pupils are to find synonyms for words (e.g. Madsen & Mohammad-Roe 2020: 233, reproduced in Appendix 1B; section 4.2.3 of this article gives a full list of textbook exercises of this type). A first difficulty of this type of textbook exercise is where to find synonyms. It is not always indicated where synonyms are to be found, but two of the book series (*Enter* and *Stages*) recommend using an online thesaurus. A second difficulty of this type of textbook exercise is that even when using a thesaurus, pupils may be unsure of which synonym to use. For example, thesaurus.com gives 45 synonyms for *nice* in the sense of 'likable, agreeable'. This difficulty is exacerbated when there is a word that can have quite different meanings: as one textbook informs pupils, 'Watch out! Make sure that you have found a synonym that fits the sentence. Example: There are many synonyms for the word *to hit*. *To hit* can mean both *to strike* and *to affect*… You cannot say *he affected the ball*' (Diskin, Winsvold & Kasbo 2020b: 74).

Exercise 1, consisting of three tasks, uses the 'Similar words' and 'Examples' features of SKELL, and the 'finds similar words' feature of Netspeak. The corpora reveal how frequently a given synonym is found in the corpus, and they provide evidence of how a synonym is used. The corpus exercise therefore does not have the two abovementioned difficulties. First, SKELL and Netspeak are sites where pupils can find synonyms. Second, the synonyms are shown in context, and from seeing evidence of the synonyms in use, pupils can become more knowledgeable about each synonym. Pupils are asked to replace *nice* in *It's a nice note, Danny* with a synonym. Thirty synonyms for *nice* can be found using the 'Similar words' feature of SKELL, but pupils can see from the resulting word cloud which of the synonyms are most commonly used. Pupils can use the 'Examples' feature of SKELL to see how each synonym is used in context. The 'finds similar words' feature of Netspeak also provides synonyms, but instead of a word cloud, the synonyms are arranged in a list by frequency. The example sentences show how each synonym is used in context. Through SKELL and Netspeak, pupils can discover which synonyms may fit the sentence. SKELL and Netspeak search different corpora, so they search different language data. The pupils are asked whether the most frequent synonym in Netspeak fits the sentence as well as the one the pupil has chosen from SKELL. All of this involves drawing conclusions from data, which requires the critical thinking described in section 3.1.

A limitation of the exercise is that the word cloud that the 'Similar words' feature of SKELL creates is different for every user, and different each time the screen is refreshed. Normally the word cloud contains only synonyms, but non-synonyms can appear. For example, in a word cloud for the other of the adjectives in the exercise, *hostile*, the antonym *friendly* has appeared.

Critical thinking will be especially necessary when this happens, obliging the same discussion about the difference between synonyms and similar words that occurs in Exercise 2.

Exercise 2. Corpus exercise for year 8 (Appendix 2B).

This corpus exercise enhances a year 8 textbook exercise asking pupils to collect examples of synonyms (Madsen & Mohammad-Roe 2020: 232, reproduced in Appendix 1C). A first difficulty of the textbook exercise is where to find synonyms, and a second difficulty is when synonyms are found, the pupils can be uncertain which to use.

Exercise 2 uses the 'Similar words' and 'Examples' features of SKELL, and the 'finds similar words' feature of Netspeak. These reveal how frequently a synonym is found in the corpus, and evidence of how a synonym is used. The corpus exercise therefore does not have the two abovementioned difficulties. Many synonyms can be found using the 'Similar words' feature of SKELL, and fewer can be found using the 'finds similar words' feature of Netspeak. The original textbook exercise asked for four synonyms of the pupil's chosen word; Netspeak will only give three synonyms of the example word *happy*, so the exercise is not worded to ask for a specific number of synonyms.

The exercise reveals a limitation of the 'Similar words' feature of SKELL: 'You will get a list of up to 40 most similar words' and these are described as 'Synonyms and similar words' (Baisa & Suchomel 2014), meaning that not all of these are synonyms. For example, thesaurus.com would not give *friendly* as a synonym for *happy*, but SKELL includes *friendly* as a similar word. The pupil now needs to learn the difference between synonyms and similar words, so there must be discussion about this, but this is a chance to exercise critical thinking. By using the 'Examples' feature and seeing each synonym in context, pupils must use critical thinking to decide what is a synonym for *happy* and what is not.

Exercise 3. Corpus exercise for year 10 (Appendix 2C).

This corpus exercise enhances a year 10 textbook exercise asking pupils to explore the meanings of 'words that can replace' *say* (i.e. synonyms for *say*) (Pettersen & Røkaas 2021b: 28, reproduced in Appendix 1D). When pupils are instructed to use the verbs 'in sentences', the difficulty is in knowing how to use each synonym.

Exercise 3 uses the textbook's twenty-synonym 'verb wheel' with SKELL. By using a corpus, the year 10 pupils get access to a large number of sentences that use the various synonyms for *say*, which gives them more data about each verb, and more context for exploring the meanings.

The other synonyms exercise types.

In section 4.2.3, other textbook synonym exercise types were identified. This article does not present corpus adaptations of these exercises, but here the following corpus-based approaches are suggested.

Synonyms exercise type 4: consider why an author did not use more common synonyms (textbook level: year 10). This exercise is for pupils to reflect on writing style, and a corpus would not necessarily be needed. However, the subject curriculum stresses 'authentic… situations' and 'authentic language models' (Norwegian Directorate for Education and Training 2019: 2, 4), and corpora are seen as sources of authentic language (e.g. Römer 2011: 210). A pupil could explore examples of particular synonyms in use, using SKELL and Netspeak, while considering this exercise.

Synonyms exercise type 5: match British English words with synonyms from other national varieties (textbook level: year 8). This type of exercise can be enhanced with evidence of geographical varieties of English from authentic situations. Corpora can provide this, but this is not a feature of SKELL or Netspeak. An online corpus that could work in this context is GloWbE. This corpus has been used in teaching, at least in teacher-corpus interaction, in upper secondary (Kavanagh 2021a: 10, 13, 15). GloWbE results are in the form of typical concordance lines, so pupils would need training to read them.

Synonyms exercise type 6. rewrite old-fashioned sentences in a modern style (textbook level: year 8). This type of exercise can be enhanced with evidence of historical varieties of English. This is not a feature of SKELL or Netspeak. This type of exercise highlights language change. Corpora can be used to trace changes in the language. Any number of exercises could be devised to compare language use over time, using COCA, and the related Corpus of Historical American English (COHA) (Davies 2010), for example. COCA has been used in both teacher-corpus and learner-corpus interaction in grades 9 and 10 (Kavanagh 2021a), but COCA (and COHA) results are in typical concordance lines which pupils would need training to read (the COCA usability issue was mentioned above in section 2.2).

Synonyms exercise type 7: use a Norwegian dictionary to see whether English translations are synonyms (textbook level: years 8-10). There is scope for investigating each of the English words in SKELL and Netspeak to see how the usage differs among the words. Something more desirable here would be to stick closely to the translation theme by using a parallel corpus. A corpus with both languages would allow pupils to compare words easily. The English-Norwegian

Parallel Corpus (https://www.hf.uio.no/ilos/english/services/knowledge-resources/omc/enpc/) might be a candidate for this, if it can be established that its texts are at a language level suitable for year 10 learners (e.g. by comparing its level with Burner, Carlsen & Steinman 2020, a book of texts for lower secondary).

Synonyms exercise type 8: use synonyms as clues in a vocabulary game called 'guess my word' (Pettersen & Røkaas 2021b: 257) (textbook level: year 10). This is not a synonyms exercise per se. The instructions suggest using synonyms as clues. In the same way, sentences from a corpus with the word missing could also be used as clues.

5.2. Collocations exercises

Exercise 4. Corpus exercise for year 10 (Appendix 2D).

This corpus exercise revises a year 10 textbook exercise (Diskin & Winsvold 2021: 155, reproduced in Appendix 1F). The textbook exercise consists of what it calls 'expressions', for example *throw a switch* and *surge of power*, which are presented to the pupil in a short text as examples of collocations. Pupils are to explain the meanings of the expressions, and consider why the use of collocations improves a text. The difficulty is that there are no other examples or evidence, only the short text.

Exercise 4 revises the textbook exercise for use without the short text. Instead the pupil is directed to evidence of the expressions in actual usage. From SKELL, pupils get more evidence of the expressions in use, in order to work out meanings from context. SKELL is particularly fruitful here because it will find examples with different forms of a verb. Results from the search term *throw a switch* will include examples like *The image of Chantal vanished instantly as though someone had thrown a switch,* so the form *thrown* is included in a search for *throw*. Some of the SKELL examples like that one, however, might seem strange to pupils out of context, so perhaps this exercise could be used with the short text as well, enhancing the textbook exercise rather than revising it.

Exercise 5. Corpus exercise for year 10 (Appendix 2E).

This corpus exercise can either replace or enhance (see below) a year 10 textbook exercise, a gap-fill exercise, 'Choose the best word to complete each collocation', where pupils choose between *strong*, *deep* and *heavy* (Pettersen & Røkaas 2021b: 226, see Appendix 1G). A potential disadvantage of this textbook exercise is that the semantic patterns are provided to the pupils in advance. The exercise explains the semantic pattern of the nouns that collocate with each adjective, for example 'The word *strong* is often used in collocations with facts and opinions and the senses.

strong argument" (ibid.). Year 10 pupils could investigate this for themselves, an activity which would create an opportunity for in-depth learning.

In Exercise 5, there are two tasks. Pupils are asked to investigate *strong, deep* and *heavy* in Netspeak, collecting evidence about the words. The aim is for the pupils to see that these adjectives collocate with different nouns. If the pupils have enough grammatical awareness to differentiate nouns from other word classes, task one can be modified to specify that nouns are sought; otherwise, the teacher will have to intervene to get the pupils to discard results where the collocate is not a noun (e.g. *a strong and* and *a strong ,*). On the other hand, pupils may be critical enough to what they find without knowing terminology (e.g. they may discount the example with the comma: *a strong ,*).

Task two of the exercise is an option influenced by the argument that collocations are not 'arbitrary' but 'motivated' (Liu & Lei 2017: 38). '[U]ncovering the motivations of collocations' (ibid.) is an alternative to learning collocations through memorization. For example, 'The reason we use *heavy* to modify an intense *rain* is that rain is made up of water; it has weight' (ibid.). Liu and Lei recommend preparing a table of typical collocations (ibid. 38-39), in which pupils may discover semantic patterns. Task two asks the pupils to create such a table themselves, but it could be made by the teacher instead, if that is too difficult.

The above description of Exercise 5 is its use as a replacement for the textbook exercise. Exercise 5 could alternatively be used to enhance the textbook exercise, by testing the collocations and semantic patterns of the three adjectives in the corpus against the collocations given to them in the textbook.

It is also possible to combine the technique of using a table of collocations with a version of the synonyms exercise that used the verb wheel (Exercise 3). The verbs in the verb wheel are synonyms, and when pupils investigate each verb, they could be asked to find collocates for it and to keep the information in a table. This would teach them the semantic context in which the verb appears (ibid.: 39), and in which they can use the verb in their own language production.

5.3. The teacher's role

In these exercises, the teacher has the guidance role conceived of as part of hands-on corpus interaction (Boulton 2012: 154). The teacher has the responsibility of introducing SKELL and Netspeak to pupils, in an instructional session. It may be beneficial for pupils to already be familiar with SKELL's most basic feature, 'Examples', and Netspeak's most basic feature, 'finds one word', beforehand. An introductory session could involve language examples pupils are

already familiar with from textbooks. For instance, in *Enter 10 Engelsk*, synonyms are introduced with the examples *arid, dehydrated* and *parched* as synonyms for *dry* (Diskin & Winsvold 2021: 201); to introduce the 'Examples' feature of SKELL, pupils could be shown how to type each of these words into the search field, and to read and discuss the example sentences for each. This gives pupils their first hands-on task, learning to search on SKELL, and to discuss the different use of each synonym. In addition, the first task in one of this article's suggested corpus exercises, Exercise 1.1 (Appendix 2A), can be seen as introductory, because it asks pupils to find synonyms for *nice*, which could be useful scaffolding for pupils introduced to synonyms through *Enter 8-10 Basic Skills,* where *nice* is one of the example words (Diskin, Winsvold & Kasbo 2020b: 74).

The teacher may also guide pupils through exercises. The suggested exercises are designed for pupils to follow step by step, and screenshots for most steps are provided. The exercises avoid including screenshots for some steps, because they would reveal in advance the results of pupils' corpus interaction. For example, a screenshot for a step near the end of Exercise 1.2 (Appendix 2A) was not provided, because it would answer a specific question that pupils are asked to answer (i.e. 'Which synonym for *nice* is used most often with *a note*?'). The final task of this exercise, Exercise 1.3 (Appendix 2A) provides no steps or screenshots, because pupils must use the same methods as the preceding tasks. The teacher can guide pupils through the same steps if necessary.

Where pupils choose which words to investigate, in Exercise 1.1 (Appendix 2A), Exercise 2 (Appendix 2B), and Exercise 3 (Appendix 2C), these exercises use examples to show pupils what to do. Where pupils are expected to discover semantic patterns for themselves, in Exercise 5.2 (Appendix 2E), the question is somewhat open, but the teacher could lead discussion, once pupils have gathered enough evidence.

Note that with these exercises, the teacher is not obliged to do as much work as in the details provided by Breyer (2009: 156) in section 1.1. First, the teacher does not have to possess more corpus literacy than being able to use the most accessible features of SKELL and Netspeak, and interpret results. Second, the teacher does not have to assess the suitability of materials in the learning context any more than with textbook exercises. Third, the teacher obviously does not need to design the exercises, but through exposure to SKELL and Netspeak, and the way these corpus exercises relate to the textbooks and the curriculum, the teacher may wish to design similarly targeted exercises, for other language topics.

6. Concluding remarks

The research question was *How can corpora be used in the English subject in lower secondary school with the current curriculum in Norway?* The suggested corpus exercises are intended to show how this can be done. They match the subject curriculum's competence aims in the same way as the corresponding textbook exercises, and also meet the digital skills aims of the subject curriculum. In terms of the core curriculum, they provide opportunities for critical thinking and in-depth learning.

The textbook exercises have difficulties or disadvantages, as explained in section 5. The suggested corpus exercises provide more examples, evidence, and context. Corpus-based teaching materials can therefore replace, enhance or revise textbook exercises.

Corpus use and DDL are not mainstream in the English subject, and obstacles to its use were enumerated above in section 1.4. The choice of corpora, SKELL and Netspeak, aimed to mitigate problems of usability, inaccessibility, expense, and the necessity for advanced prior digital skills. The curriculum itself encourages the use of language materials and explicit learning of language, mitigating other obstacles mentioned.

The suggested exercises covered only two vocabulary language elements. Corpus exercises can also be designed for grammar language elements, and other curricular competence aims can be met. To give a single grammar example, modality can be linked to year 10 competence aims (Flognfeldt & Lund 2021: 204), and this is reflected in exercises with modals in *Engelsk 10 fra Cappelen Damm Student's Book, Enter 9 Engelsk* and *Enter 8-10 Basic Skills*. There is previous literature to draw upon when designing such exercises, for example, Papaioannou, Mattheoudakis and Agathopoulou (2020) write about DDL work with modals for 15-16-year-olds.

There has long been a difficulty in connecting corpus linguistics to teaching practice. Materials development could be a fruitful approach, especially if teachers cannot become confident corpus linguists themselves. Past attempts to connect corpora and teaching practice have focused on educating teachers and teacher-education students in corpus linguistics. This article attempts another method: suggesting corpus exercises that teachers may find accessible, and that relate to the curriculum that teachers are familiar with and obliged to work with. Researchers could devote time to designing hands-on step-by-step exercises, or other corpus-based teaching materials, for specific curricula, which can then be made available to teachers, perhaps working in collaboration with teachers, to ensure exercises and materials fit not only the curriculum but other needs perceived by teachers, and classroom focus.

References

Baisa, V., and Suchomel, V. (2014), 'SKELL—web interface for English language learning', in A. Horák and P. Rychlý (eds.), *Proceedings of RASLAN 2014 (recent advances in Slavonic natural language processing)* (NLP Consulting, Czech Republic), 63–70. Details of the English Corpus for SKELL: <https://www.sketchengine.eu/english-skell-corpus/>, last accessed 28 August 2023. Information about the SKELL 'Similar words' feature: <https://www.sketchengine.eu/skell/#tAb-id-3>, last accessed 28 August 2023.

Boulton, A. (2009), 'Testing the Limits of Data-Driven Learning: Language Proficiency and Training', *ReCALL*, 21/1: 37-54.

Boulton, A. (2012), 'Hands-on/hands-off: Alternative approaches to data-driven learning', in J. Thomas and A. Boulton (eds.), *Input, Process and Product: Developments in Teaching and Language Corpora* (Masaryk University Press; Czech Republic), 152-168. < https://hal.science/hal-00503034>, last accessed 28 August 2023.

Boulton, A., and Cobb, T. (2017), 'Corpus use in language learning: A meta-analysis', *Language Learning*, 67/2: 348-393.

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.

Breyer, Y. (2009), 'Learning and teaching with corpora: reflections by student teachers', *CALL*, 22/2: 153-172.

Breyer, Y. (2011), *Corpora in Language Teaching and Learning: Potential, Evaluation, Challenges* (Peter Lang GmbH Internationaler Verlag der Wissenschaften; Berlin) <https://www.researchgate.net/publication/264499058_Corpora_in_Language_Teaching_and_Learning_Potential_Evaluation_Challenges>, last accessed 28 August 2023.

Burner, T. (2020), 'What the 2020 curriculum reform means for English teachers',

in C. Carlsen, M. Dypedahl, and S.H. Iversen (eds.), *Teaching and Learning English* (2nd edn., Cappelen Damm Akademisk; Oslo), 53-61.

Burner, T., Carlsen, C., and Steinman, K. (eds). (2020), *Echo: Texts in English 8-10* (Fagbokforlaget; Bergen).

Callies, M. (2019), 'Integrating corpus literacy into language teacher education: the case of learner corpora', in S. Götz and J. Mukherjee (eds.), *Learner Corpora and Language Teaching* (John Benjamins Publishing Company; Amsterdam), 245-263.

Cobb, T. (1997), 'Is there any measurable learning from hands-on concordancing?', *System*, 25/3: 301-315.

Crosthwaite, P. (ed.). (2020), *Data-driven learning for the next generation: Corpora and DDL for Pre-tertiary learners* (Routledge; London).

Davies, M. (2002-), English-Corpora.org <https://www.english-corpora.org/>.

Davies, M. (2008-), Corpus of Contemporary American English (COCA): <https://www.english-corpora.org/coca/>.

Davies, Mark. (2010), Corpus of Historical American English (COHA) <https://www.english-corpora.org/coha/>.

Davies, M. (2013), Corpus of Global Web-Based English <https://www.english-corpora.org/glowbe/>.

Diskin, E., Kasbo, K., and Winsvold, K.G. (2020a), *Enter 8 Engelsk* (Gyldendal; Oslo).

Diskin, E., and Winsvold, K.G. (2020), *Enter 9 Engelsk* (Gyldendal; Oslo).

Diskin, E., and Winsvold, K.G. (2021), *Enter 10 Engelsk* (Gyldendal; Oslo).

Diskin, E., Winsvold, K.G., and Kasbo, K. (2020b), *Enter 8-10 Basic Skills* (Gyldendal; Oslo).

Ebrahimi, A., and Faghih, E. (2016), 'Integrating corpus linguistics into online language teacher education programs', *ReCALL*, 29/1: 120–135. doi:10.1017/S0958344016000070.

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. doi:10.1080/09571736.2010.520728.

Farr, F. and Karlsen, P.H. (2022), 'DDL Pedagogy, Participants, and Perspectives', in R.R. Jablonkai and E. Csomay (eds.), *The Routledge Handbook of Corpora and English Language Teaching and Learning* (Routledge; London and New York), 329-343.

Flognfeldt, M., and Lund, R. (2021), *English for Teachers and Learners* (2nd edition, Cappelen Damm Akademisk; Oslo).

Gilje, Ø., Ingulfsen, L., Dolonen, J.A., Furberg, A., Rasmussen, I., Kluge, A., Knain, E., Mørch, A., Naalsund, M., and Skarpaas, K.G. (2016), *Med ARK&APP – Bruk av læremidler og ressurser for læring på tvers av arbeidsformer* (University of Oslo; Oslo).

Gilquin, G. and Granger, S. (2022), 'Using data-driven learning in language teaching', in A. O'Keeffe and M.J. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics* (2nd edn., Routledge; London), 430-442.

Haegi, E., Madsen, T., and Mohammad-Roe, S. (2020), *Engelsk 9 fra Cappelen Damm Student's Book* (Cappelen Damm; Oslo).

Haegi, E., Madsen, T., and Mohammad-Roe, S. (2021), *Engelsk 10 fra Cappelen Damm Student's Book* (Cappelen Damm; Oslo).

Johns, T. (1991), 'Should You Be Persuaded – Two Samples of Data-driven Learning Materials', *ELR Journal Classroom Concordancing*, 4: 1-16.

Karlsen, P.H. (in preparation), 'Integrating Multimedia Corpora in the Secondary School Classroom in Norway'.

Karlsen, P.H., and 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.

Kavanagh, B. (2021a), '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.

Kavanagh, B. (2021b), 'Norwegian In-service Teachers' Perspectives on Language Corpora in Teaching English', *Nordic Journal of Language Teaching and Learning*, 9/2: 90-106.

Koellner, K, and Greenblatt, D. (2018), 'Inservice Teacher Education' entry in Oxford Bibliographies <https://www.oxfordbibliographies.com/view/document/obo-9780199756810/obo-9780199756810-0196.xml>, last accessed 28 August 2023.

Lee, H., Warschauer, M., and Lee, J.H. (2019), 'The Effects of Corpus Use on Second Language Vocabulary Learning: A Multilevel Meta-analysis', *Applied Linguistics*, 40/5: 721-753.

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. doi:10.1017/S095834401400010X.

Leńko-Szymańska, A. (2017), 'Training teachers in data-driven learning: tackling the challenge', *Language Learning and Technology*, 21/3: 217-241.

Leńko-Szymańska, A., and Boulton, A. (2015), 'Data-driven learning in language pedagogy', in A. Leńko-Szymańska and A. Boulton (eds.), *Multiple Affordances of Language Corpora for Data-driven Learning* (John Benjamins Publishing Company, Amsterdam and Philadelphia), ix-xxii.

Liu, D., and Lei, L. (2017), *Using Corpora for Language Learning and Teaching* (Tesol Press; Alexandria).

McEnery, T., and Hardie, A. (2012), *Corpus Linguistics: Method, Theory and Practice* (Cambridge University Press; Cambridge).

McEnery, T., and 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)* (Routledge: London and New York), 364-380.

Madsen, T., and Mohammad-Roe, S. (2020), *Engelsk 8 fra Cappelen Damm Student's Book* (Cappelen Damm; Oslo).

Meunier, F. (2022), 'Revamping DDL: Affordances of Digital Technology', in R.R. Jablonkai and E. Csomay (eds.), *The Routledge Handbook of Corpora and English Language Teaching and Learning* (Routledge; London and New York), 344-360.

Mukherjee, J. (2004), 'Bridging the gap between applied corpus linguistics and the reality of English language teaching in Germany', in U. Connor and T.A. Upton (eds.), *Applied corpus Linguistics: A Multidimensional perspective* (Rodopi; Amsterdam and New York), 239-250.

Norges Offentlige Utredninger. (2015), *NOU 2015: 8. Fremtidens Skole – Fornyelse av fag og kompetanser* <https://www.regjeringen.no/no/dokumenter/nou-2015-8/id2417001>, last accessed 4 July 2022.

Norwegian Directorate for Education and Training. (2013), *English Subject Curriculum* <https://data.udir.no/kl06/ENG1-03.pdf?lang=eng>, last accessed 23 September 2022.

Norwegian Directorate for Education and Training. (2017), *Core curriculum – values and principles for primary and secondary education* <https://www.udir.no/lk20/overordnet-del/?lang=eng>, last accessed 4 July 2022.

Norwegian Directorate for Education and Training. (2019), *Curriculum in English* <https://data.udir.no/kl06/v201906/laereplaner-lk20/ENG01-04.pdf?lang=eng>, last accessed 4 July 2022.

Papaioannou, V., Mattheoudakis, M., and Agathopoulou, E. (2020), 'Data-driven learning in a Greek secondary education setting: the implementation of a blended approach', in P. Crosthwaite, Peter (ed.), *Data-driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary learners* (Routledge; London), 187-207.

Pettersen, S., and Røkaas, F. (2020), *Stages 8 Engelsk for ungdomstrinnet* (Aschehoug Undervisning; Oslo).

Pettersen, S., and Røkaas, F. (2021a), *Stages 9 Engelsk for ungdomstrinnet* (Aschehoug Undervisning; Oslo).

Pettersen, S., and Røkaas, F. (2021b), *Stages 10 Engelsk for ungdomstrinnet* (Aschehoug Undervisning; Oslo).

Pinto, P.T., Crosthwaite, P., de Carvalho, C.T., Spinelli, F., Serpa, T., Garcia, W., and Ottaiano, A.O. (2023) (eds.), *Using Language Data to Learn About Language: A Teachers' Guide to Classroom Corpus Use* (University of Queensland; Brisbane). doi:10.14264/3bbe92d.

Römer, U. (2011), 'Corpus research applications in second language teaching', *Annual Review of Applied Linguistics* 31: 205-225. doi:10.1017/S0267190511000055.

Skulstad, A.S. (2020), 'Communicative competence', in A. Fenner and A.S. Skulstad (eds.), T*eaching English in the 21st Century: Central Issues in English Didactics* (2nd edn., Fagbokforlaget; Bergen), 43-68.

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.

Vyatkina, N. (2016), 'Data-driven Learning of Collocations: Learner Performance, Proficiency, and Perceptions', *Language Learning & Technology*, 20/3: 159-179.

Zareva, A. (2017), 'Incorporating corpus literacy skills into TESOL teacher training', *ELT Journal*, 71/1: 69-79. doi:10.1093/elt/ccw045.

Appendix 1. Textbook Excerpts and Exercises.

Contents:

1A. How synonyms are introduced to pupils in textbooks: two examples

synonyms – ord som betyr det samme, f.eks. *large* og *big*
thesaurus – synonymordbok
replace – erstatte

Synonyms

Synonyms are words that mean almost the same. The words in this word cloud are some of the most common English words. To make your language more varied and interesting, you should try to find synonyms for these common words.

Common words	Synonyms
nice	charming, lovely, sweet, good-natured, pleasant
bad	awful, unacceptable, evil, rotten, wicked
happy	delighted, overjoyed, pleased, satisfied, content

He struck the ball.

How to find synonyms?

- Find a thesaurus on the internet. A thesaurus is almost like a dictionary, but it gives synonyms for words.
- Look up the Norwegian word in a dictionary. You will find different English synonyms for the word.

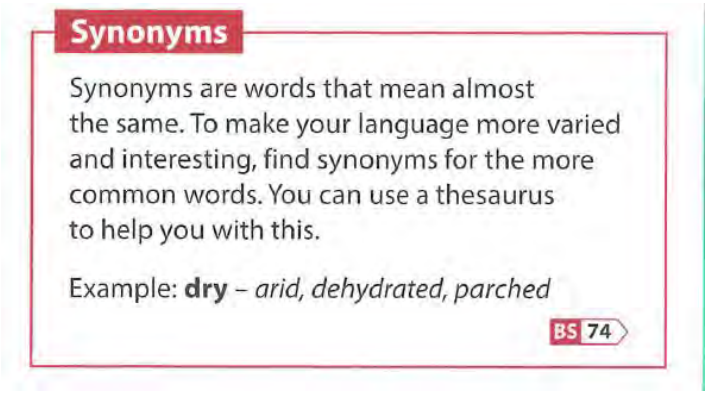
Watch out!

- Make sure that you have found a synonym that fits the sentence.
  Example: There are many synonyms for the word *to hit*.
  To hit can mean both *to strike* and *to affect*.

  Example 1: *He hit the ball.*
  You can replace *hit* with *strike*: *he struck the ball.*
  You cannot say *he affected the ball.*

  Example 2: *Her grandfather's death hit her hard.*
  You can replace *hit* with *affect*: *her grandfather's death affected her.*
  You cannot say *her grandfather's death struck her.*

From: *Enter 8-10 Basic Skills*, for years 8-10 (Diskin, Winsvold & Kasbo 2020b: 74).[13]



Synonyms

Synonyms are words that mean almost the same. To make your language more varied and interesting, find synonyms for the more common words. You can use a thesaurus to help you with this.

Example: **dry** – *arid, dehydrated, parched*

BS 74

From: *Enter 10 Engelsk*, a textbook for year 10 (Diskin & Winsvold 2021: 201).

[13] Gyldendal Norsk Forlag AS granted permission for the reproductions from the *Enter* book series.

1B. Synonyms. Textbook exercise, first example.

b) Write as many synonyms as you can for each of the following words:

say, walk, think, smile

From: *Engelsk 8 fra Cappelen Damm Student's Book*, a textbook for year 8 (Madsen & Mohammad-Roe 2020: 233).[14]

1C. Synonyms. Textbook exercise, second example.

Language and grammar

a) What is a synonym? Write down a definition. Then, choose a word and write at least four examples of synonyms for that word.

From: *Engelsk 8 fra Cappelen Damm Student's Book*, a textbook for year 8 (Madsen & Mohammad-Roe 2020: 232).

[14] Cappelen Damm AS granted permission for the reproductions from the *Engelsk* book series.

1D. Synonyms. Textbook exercise, third example.

VOCABULARY

5 Synonyms

a Explore how the meaning of the sentences below changes if you replace "said" with "hissed" or "grunted".
"No money from your dad again, Blake," Mum said.
"We'll be all right, Mum," Blake said.
"I haven't come here to hurt you, Blake," Mr Vilman said.

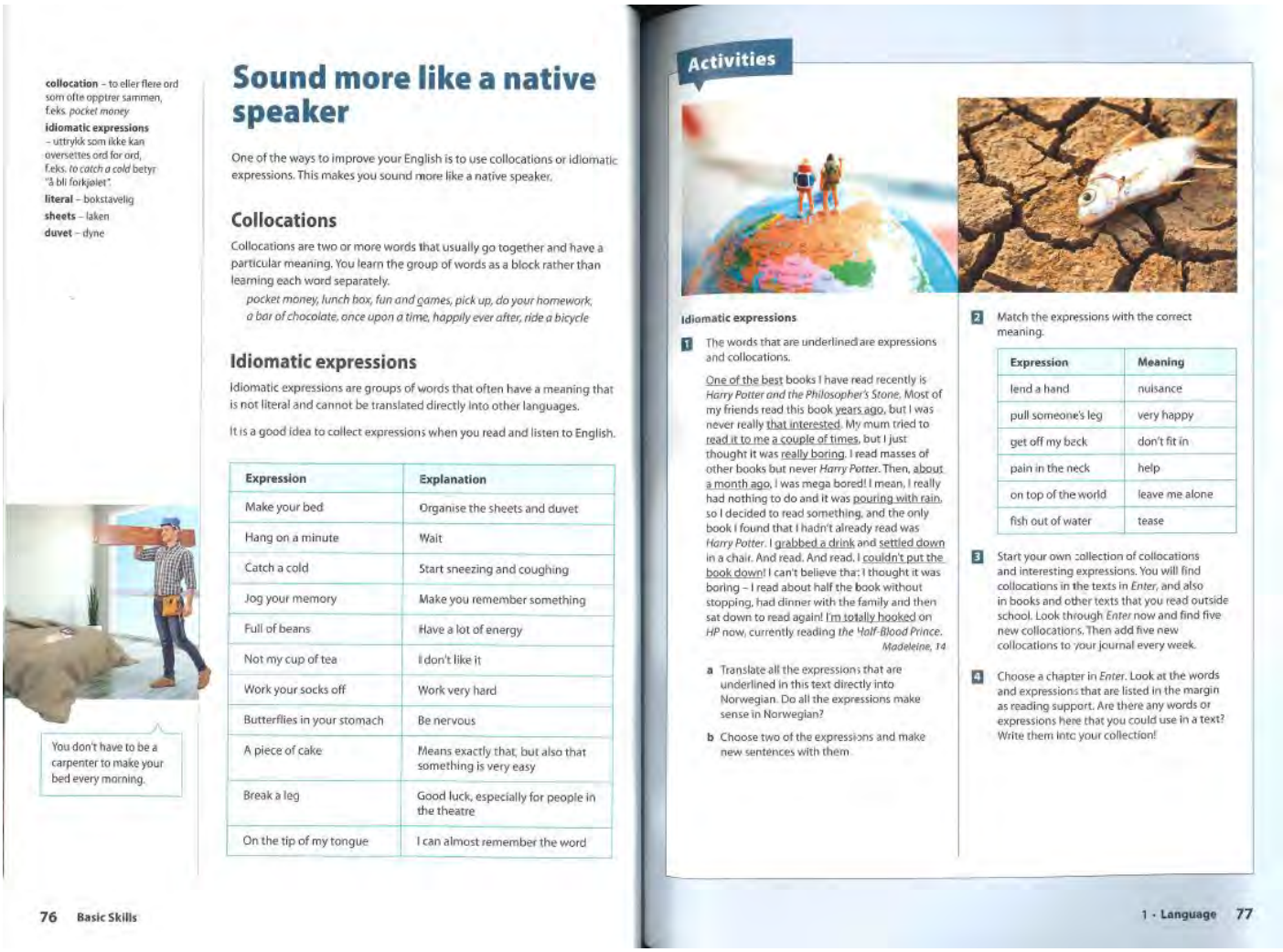
b Use the verb wheel below and choose six other words that can replace "said". Use them in sentences. Discuss with a partner if the synonym for "said" gives the sentence a certain feeling. Does it makes the sentence angry, thoughtful, hopeful, sad, etc?

From: *Stages 10 Engelsk for ungdomstrinnet*, a textbook for year 10 (Pettersen & Røkaas 2021b: 28).[15]

Part (a) refers to a short story not reproduced here, 'The Offer' by Sophie McKenzie.

[15] H. Aschehoug & Co. (W. Nygaard) AS granted permission for the reproductions from the *Stages* book series; Markus Fogth-Jakobsen granted permission for the reproduction of the verb wheel illustration.

1E. How collocations are introduced to pupils in textbooks: two examples



collocation – to eller flere ord som ofte opptrer sammen, f.eks. *pocket money*
idiomatic expressions – uttrykk som ikke kan oversettes ord for ord, f.eks. *to catch a cold* betyr "å bli forkjølet".
literal – bokstavelig
sheets – laken
duvet – dyne

Sound more like a native speaker

One of the ways to improve your English is to use collocations or idiomatic expressions. This makes you sound more like a native speaker.

Collocations

Collocations are two or more words that usually go together and have a particular meaning. You learn the group of words as a block rather than learning each word separately.

pocket money, lunch box, fun and games, pick up, do your homework, a bar of chocolate, once upon a time, happily ever after, ride a bicycle

Idiomatic expressions

Idiomatic expressions are groups of words that often have a meaning that is not literal and cannot be translated directly into other languages.

It is a good idea to collect expressions when you read and listen to English.

Expression	Explanation
Make your bed	Organise the sheets and duvet
Hang on a minute	Wait
Catch a cold	Start sneezing and coughing
Jog your memory	Make you remember something
Full of beans	Have a lot of energy
Not my cup of tea	I don't like it
Work your socks off	Work very hard
Butterflies in your stomach	Be nervous
A piece of cake	Means exactly that, but also that something is very easy
Break a leg	Good luck, especially for people in the theatre
On the tip of my tongue	I can almost remember the word

You don't have to be a carpenter to make your bed every morning.

76 Basic Skills

Activities

Idiomatic expressions

1 The words that are underlined are expressions and collocations.

One of the best books I have read recently is *Harry Potter and the Philosopher's Stone*. Most of my friends read this book years ago, but I was never really that interested. My mum tried to read it to me a couple of times, but I just thought it was really boring. I read masses of other books but never *Harry Potter*. Then, about a month ago, I was mega bored! I mean, I really had nothing to do and it was pouring with rain, so I decided to read something, and the only book I found that I hadn't already read was *Harry Potter*. I grabbed a drink and settled down in a chair. And read. And read. I couldn't put the book down! I can't believe that: I thought it was boring – I read about half the book without stopping, had dinner with the family and then sat down to read again! I'm totally hooked on *HP* now, currently reading *the Half-Blood Prince*.

Madeleine, 14

a Translate all the expressions that are underlined in this text directly into Norwegian. Do all the expressions make sense in Norwegian?

b Choose two of the expressions and make new sentences with them.

2 Match the expressions with the correct meaning.

Expression	Meaning
lend a hand	nuisance
pull someone's leg	very happy
get off my back	don't fit in
pain in the neck	help
on top of the world	leave me alone
fish out of water	tease

3 Start your own collection of collocations and interesting expressions. You will find collocations in the texts in *Enter*, and also in books and other texts that you read outside school. Look through *Enter* now and find five new collocations. Then add five new collocations to your journal every week.

4 Choose a chapter in *Enter*. Look at the words and expressions that are listed in the margin as reading support. Are there any words or expressions here that you could use in a text? Write them into your collection!

1 · Language 77

From: *Enter 8-10 Basic Skills* (Diskin, Winsvold & Kasbo 2020b: 76). Here collocations are introduced alongside idiomatic expressions.

9 Collocations

A collocation is two or more words that go together naturally. It is important to learn collocations because they make your English sound fluent and natural. Here are some common English collocations with the words *strong*, *deep* and *heavy*. Study them and then do the task.

STRONG

The word strong is often used in collocations with facts and opinions and the senses.

strong argument
strong evidence
strong feeling
strong opinion
a strong smell/taste

DEEP

The word deep is used for some strong feelings and in some expressions.

deep depression
deep devotion
in deep thought
in deep trouble
in a deep sleep (when a person won't wake up easily)

HEAVY

The word heavy is used for some weather conditions, for people with bad habits and to describe traffic.

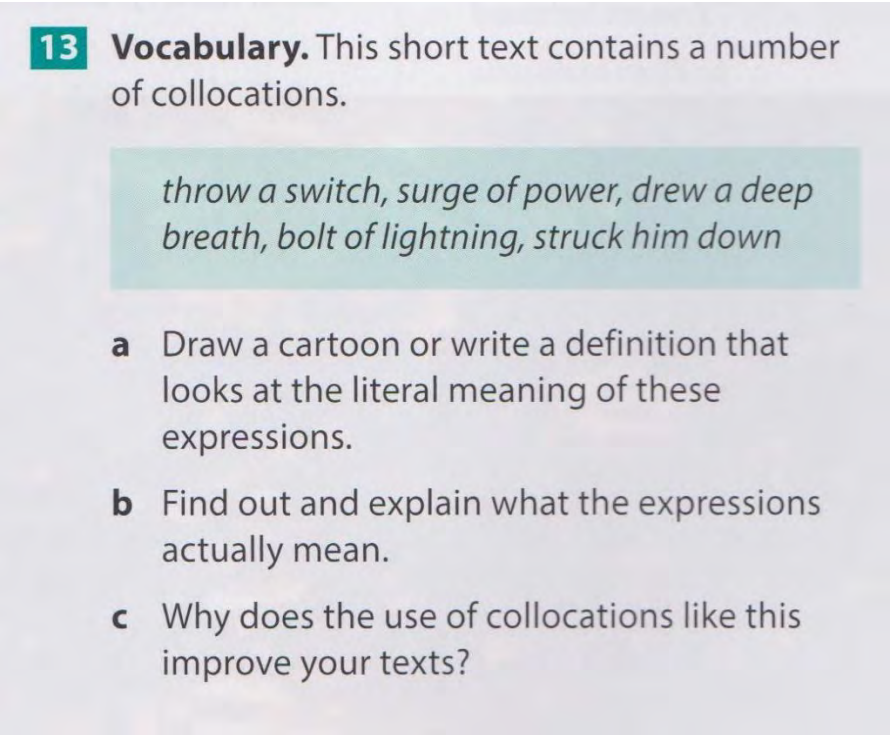
heavy rain/snow/fog
a heavy drinker
a heavy smoker
a heavy drug user
heavy traffic

Choose the best word to complete each collocation: heavy, deep or strong.

a The *heavy* traffic made us late for the appointment.
b That cheese has a ___ smell, but the taste is mild.
c After his wife died, he fell into a ___ depression.
d She was in such a ___ sleep that we couldn't wake her.
e ___ fog made it impossible for the airplane to land.
f I have a ___ feeling that I have been here before.
g If Mom finds out, we will be in ___ trouble.
h My grandfather was a ___ smoker for many years.
i There is ___ evidence linking sun exposure to skin cancer.

From: *Stages 10 Engelsk for ungdomstrinnet*, a textbook for year 10 (Pettersen & Røkaas 2021b: 226). Here collocations are introduced as part of an exercise.

1F. Collocations. Textbook exercise, first example.



13 Vocabulary. This short text contains a number of collocations.

throw a switch, surge of power, drew a deep breath, bolt of lightning, struck him down

a Draw a cartoon or write a definition that looks at the literal meaning of these expressions.

b Find out and explain what the expressions actually mean.

c Why does the use of collocations like this improve your texts?

From: *Enter 10 Engelsk*, a textbook for year 10 (Diskin & Winsvold 2021: 155). It is preceded by a short text in which the expressions are used. The short text, 'Answer' by Frederic Brown, is not reproduced here.

1G. Collocations. Textbook exercise, second example.

The second textbook exercise example is the *Stages 10 Engelsk for ungdomstrinnet* exercise, reproduced in 1E above (Pettersen & Røkaas 2021b: 226).

Appendix 2. Corpus Exercises: Using SKELL and Netspeak to Teach Synonyms and Collocations

Contents:

2A. Exercise 1. Synonyms. Corpus exercise for years 8-10.

I said to one boy, Did your mother really write this note, Danny?

He was defensive, hostile. Yeah, my mother wrote it.

It's a nice note, Danny. She writes well. [16]

The aim of this exercise is to replace two of the adjectives in this text – *hostile* and *nice* – with synonyms.

[16] Text from an adjectives and adverbs exercise in *Engelsk 10 fra Cappelen Damm Student's Book* (Haegi, Madsen & Mohammad-Roe 2021: 44).

There are three tasks to complete in this exercise. In Task One, you will use SKELL. In Task Two, you will use Netspeak. In Task Three, you can use either SKELL or Netspeak.[17]

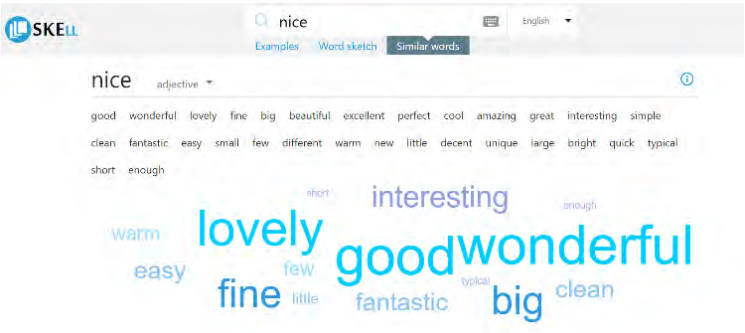
Exercise 1.1. Task One: Using SKELL.

Find synonyms for the adjective *nice* in SKELL: https://skell.sketchengine.eu/#home?lang=en

Open SKELL. In the search field, type *nice*.

Choose the 'Similar Words' feature of SKELL.

There are many synonyms for *nice*:

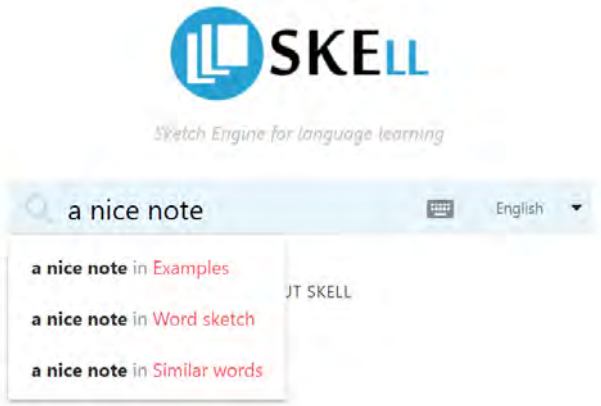


We want a synonym for *nice* that fits well in the sentence *It's a nice note, Danny*. Which of the synonyms from the word cloud fit best, do you think?

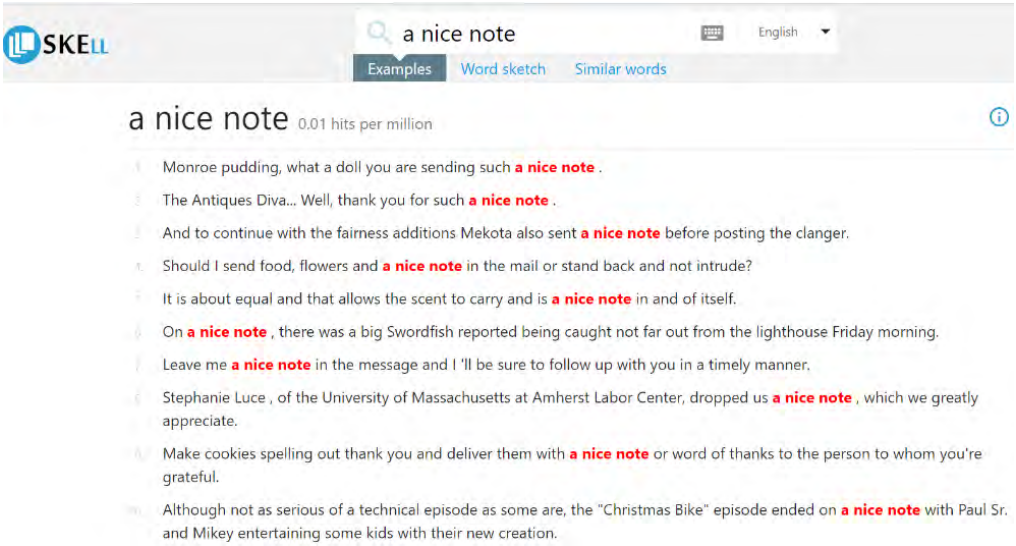
Now, click on the 'examples' tab:

[17] 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 (https://netspeak.org/) searches Google Books.

Search for: *a nice note*



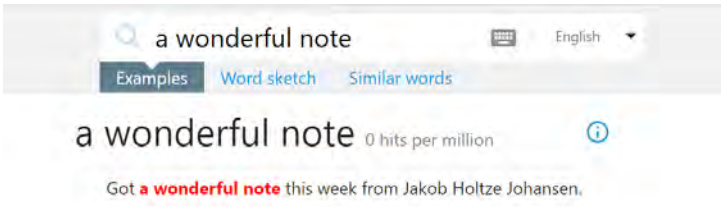
Read the results for *a nice note*.



You can see that you have found example sentences with *a nice note* in it.

Remember that there were many synonyms for *nice*.

Search for some of these synonyms with *a note*. For example, search for *a wonderful note*

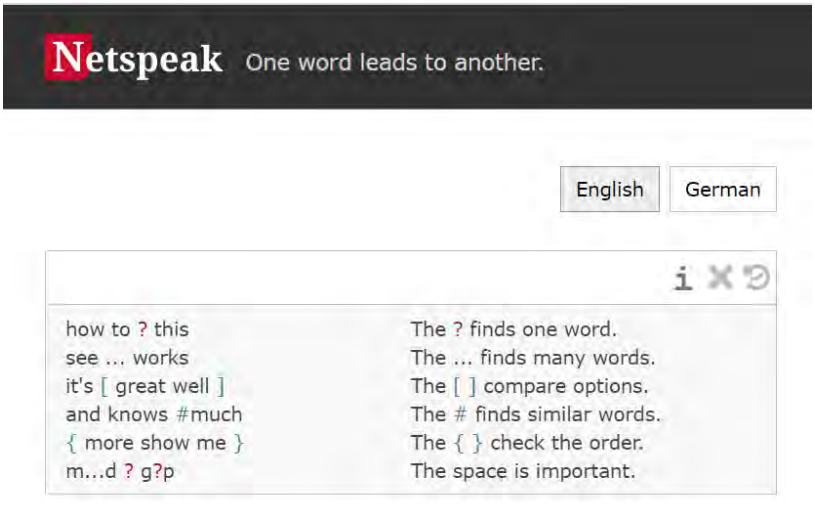


Read the examples of each synonym with *a note*.

Which synonym for *nice* best fits the sentence *It's a nice note, Danny*, do you think?

Exercise 1.2. Task Two: Using Netspeak.

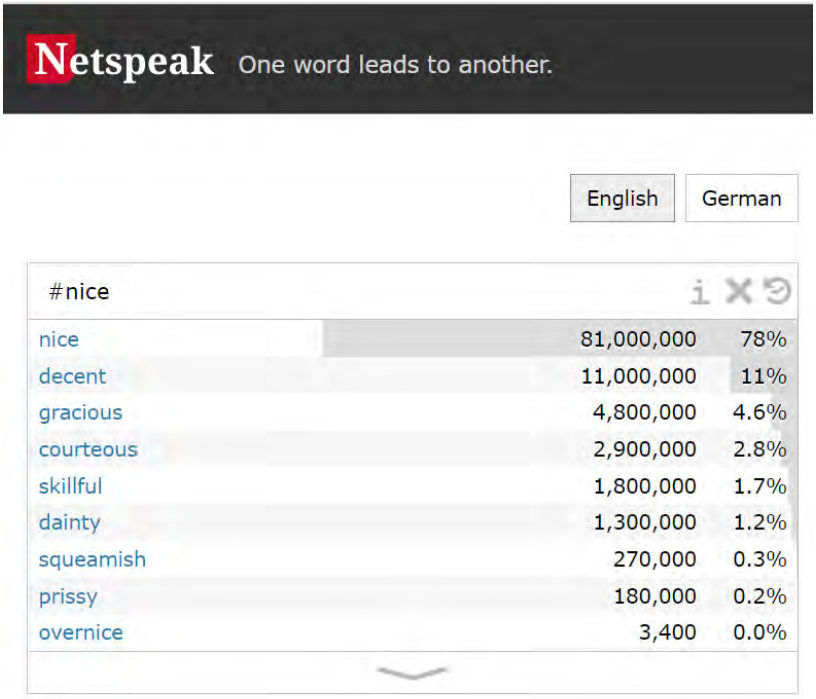
Open Netspeak: https://netspeak.org/



You use the # ('finds similar words') to find synonyms in Netspeak.

Find synonyms for *nice* by typing *#nice* into the search field.

The synonyms appear highlighted in blue:



Our aim in this overall exercise has been to replace *nice* in *It's a nice note, Danny* with a synonym. So, look at *a nice note* in Netspeak.

In the search field, type *a #nice note.*

Read the results. Which synonym for *nice* is used most often with *a note*?

Click on a synonym (highlighted in blue) to read example sentences.

Does the synonym that is used most often fit the sentence *It's a nice note, Danny* as well as the one you chose in Task One, the SKELL task?

Exercise 1.3. Task Three: Using either SKELL or Netspeak.

Find synonyms for the adjective *hostile,* using either SKELL or Netspeak.

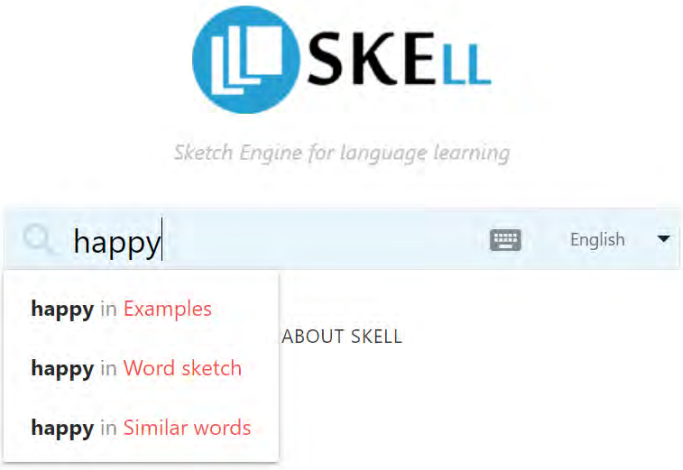
Decide which synonym for *hostile* best fits the sentence *He was defensive, hostile.*

2B. Exercise 2. Synonyms. Corpus exercise for year 8.

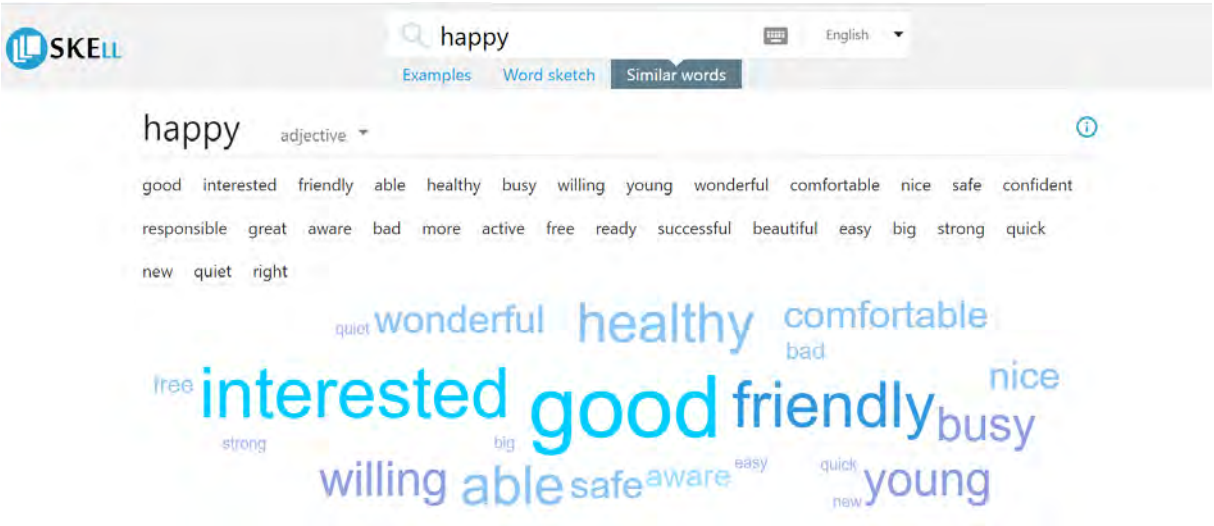
Choose a word and collect examples of synonyms of that word.

The first step is to think of a word. For example, *happy*.

Choose the 'Similar words' feature of SKELL: https://skell.sketchengine.eu/#home?lang=en

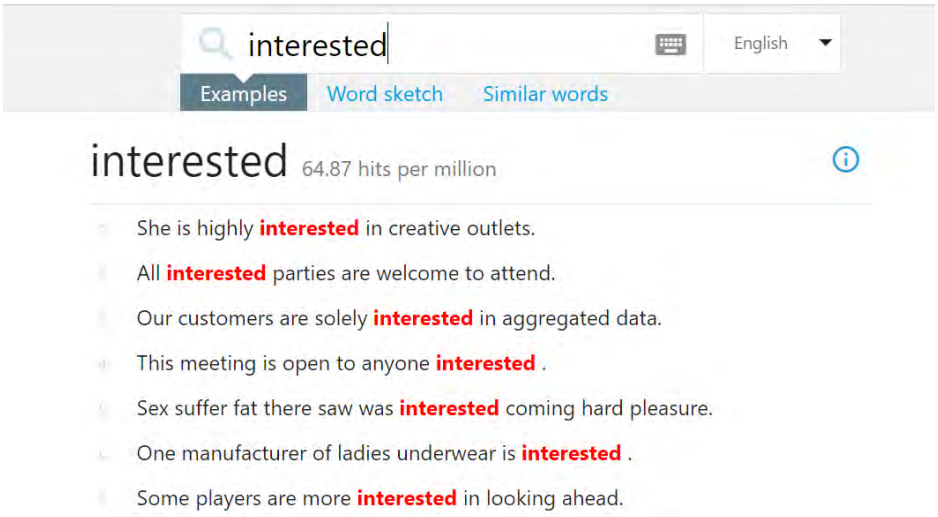


There are many similar words to *happy*:

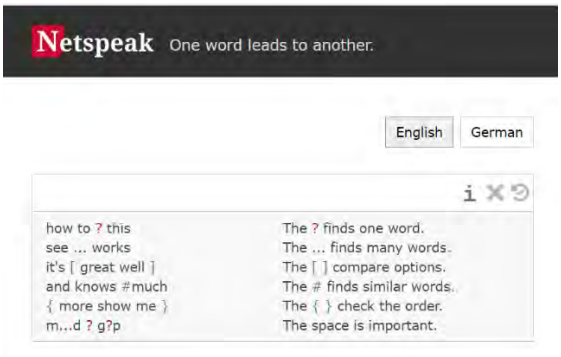


Are each of these similar words synonyms for *happy*? Are some of these words similar to *happy*, but not synonyms for *happy*?

If you are not sure how to answer, read the example sentences for each of these words in it, using the 'Examples' tab of SKELL, like *interested*:



You can also use # ('finds similar words') in Netspeak to find synonyms.



For example, to find synonyms for *happy,* search for *#happy*.

Which synonym for *happy* is used most often?

Now you have investigated *happy*. Try other words, of your choice.

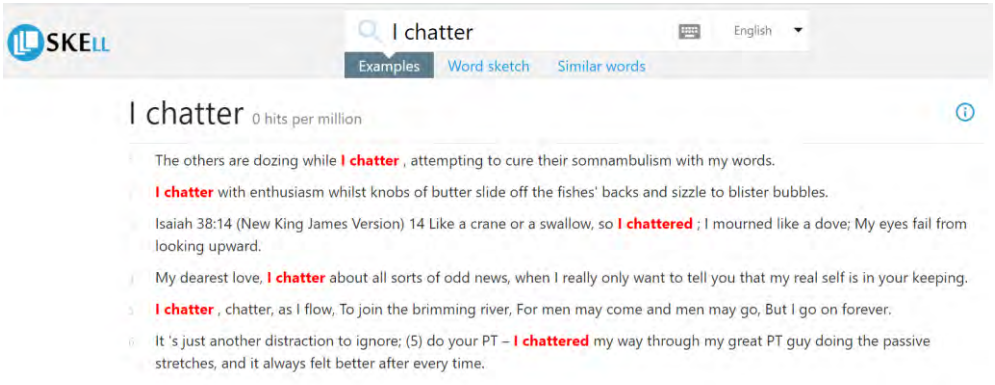
2C. Exercise 3. Synonyms. Corpus exercise for year 10.

Look at the verb wheel (Appendix 1D).

Alone or with a partner, choose words that can replace *say*.

Explore the meaning of your chosen words in SKELL.

For example, if you chose *chatter,* search for examples of *chatter* using the 'Examples' feature of SKELL. You want examples of *chatter* the verb, so search for it by using a subject, e.g. *I chatter* or *you chatter*, so that your examples will be verbs:

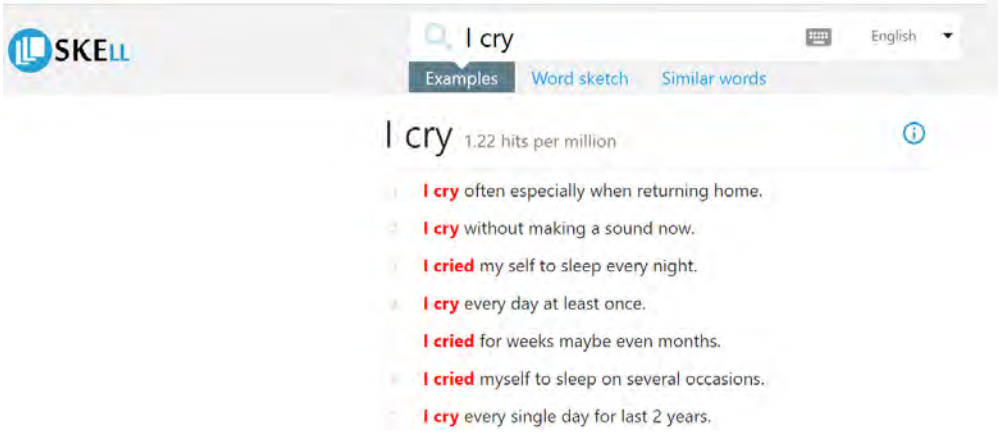


Is *I chatter* used in sentences that are very different from *I cry, I shout, I mumble,* and so on?

What happens when you replace *I chatter* with *I cry* in one of the above sentences*?*

What happens when you replace *I cry* with *I chatter* in one of the sentences below?

Chatter and *cry* are synonyms of *say*, but are they synonyms of each other?



2D. Exercise 4. Collocations. Corpus exercise for year 10.

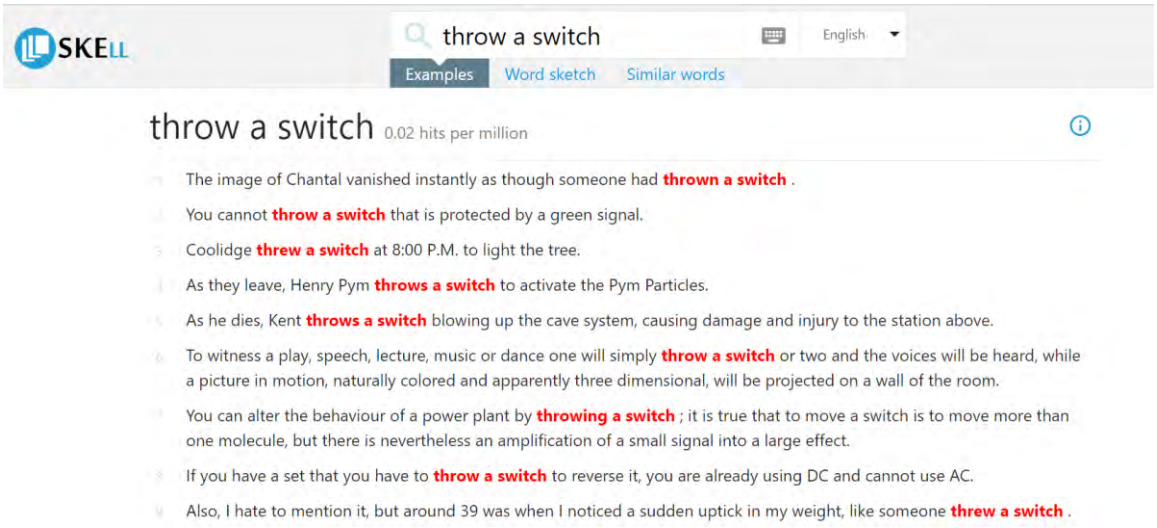
Here are five examples of collocations. In each case, the words go together to make a particular meaning:

throw a switch, surge of power, drew a deep breath, bolt of lightning, struck him down[18]

Find out what these expressions mean, from the contexts in which they are used. To do this, open SKELL: https://skell.sketchengine.eu/#home?lang=en

Search SKELL, using the 'Examples' tab.

In the search field type, for instance, *throw a switch*, and you will get these results:



Can you explain what each of the five expressions means?

[18] From Diskin & Winsvold 2021: 155.

2E. Exercise 5. Collocations. Corpus exercise for year 10.

Exercise 5.1. Task One: Using Netspeak.

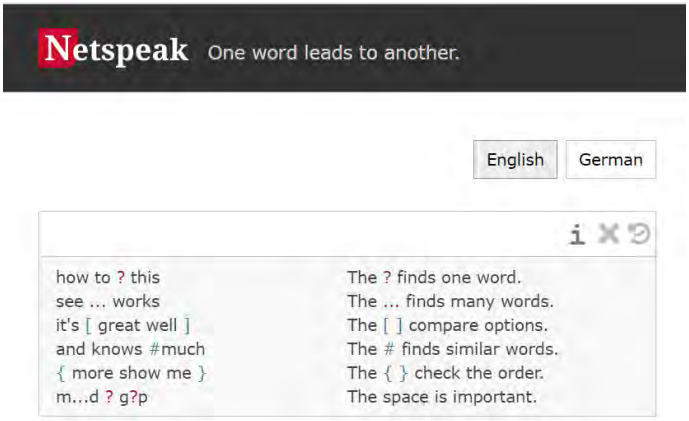
What kind of words go together with *strong*?

What kind of words go together with *deep*?

What kind of words go together with *heavy?*

To answer this, use Netspeak.

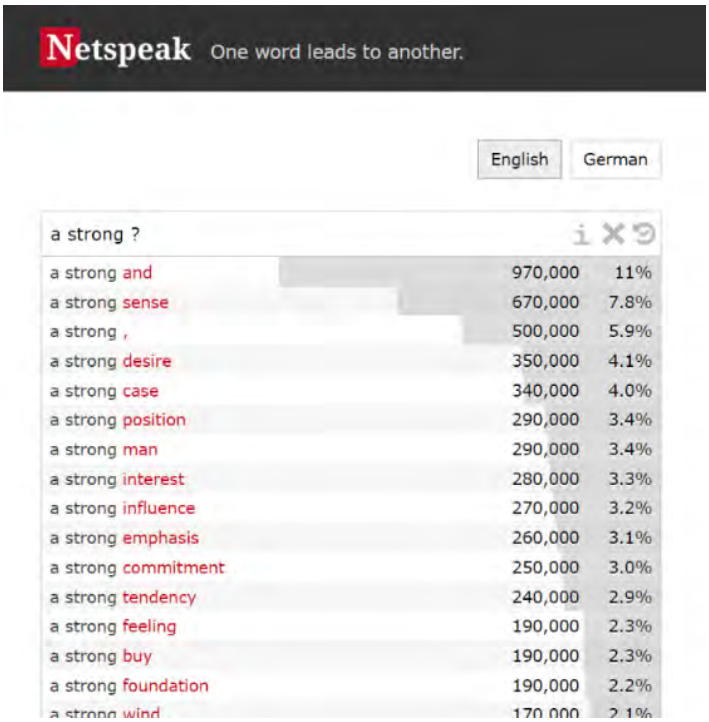
Open Netspeak: https://netspeak.org/



The question mark feature of Netspeak 'finds one word'.

Use the question mark to discover the collocations.

For example, type into the search field: *a strong ?*



Do you see what kind of things or people *strong* describes?

Now, what kind of things does the word *deep* describe?

What kind of things does the word *heavy* describe?

Exercise 5.2. Task Two: Table of collocations.

Your teacher or textbook has explained what a 'collocation' is.

Fill out this table for the common collocations of *strong, deep* and *heavy*.

Strong	**Deep**	**Heavy**
A strong sense	*A deep breath*	*A heavy burden*
A strong desire	*A deep sense*	*A heavy heart*
A strong _____	*A deep* _____	*A heavy* _____
A strong _____	*A deep* _____	*A heavy* _____
A strong _____	*A deep* _____	*A heavy* _____
A strong _____	*A deep* _____	*A heavy* _____
A strong _____	*A deep* _____	*A heavy* _____
A strong _____	*A deep* _____	*A heavy* _____
A strong _____	*A deep* _____	*A heavy* _____

Examine the collocations closely. Pay attention to the meanings of *strong, deep* and *heavy*.

Is there any pattern of meaning in the words that collocate with *strong*?

Is there any pattern of meaning in the words that collocate with *deep*?

Is there any pattern of meaning in the words that collocate with *heavy*?

Dissertation appendices

Appendix 1. NSD evaluation

Vurdering fra NSD Personvernombudet for forskning § 31

Assessment of processing of personal data

Barry Kavanagh

2418 ELVERUM

Vår dato: 11.01.2018 | Vår ref: 57613 / 3 / EPA | Deres dato: | Deres ref:

Vurdering fra NSD Personvernombudet for forskning § 31

Personvernombudet for forskning viser til meldeskjema mottatt 06.12.2017 for prosjektet:

57613	*Questionnaire for English Teachers*
Behandlingsansvarlig	*Høgskolen i Innlandet, ved institusjonens øverste leder Barry*
Daglig ansvarlig	*Kavanagh*

Vurdering

Etter gjennomgang av opplysningene i meldeskjemaet og øvrig dokumentasjon finner vi at prosjektet er meldepliktig og at personopplysningene som blir samlet inn i dette prosjektet er regulert av personopplysningsloven § 31. På den neste siden er vår vurdering av prosjektopplegget slik det er meldt til oss. Du kan nå gå i gang med å behandle personopplysninger.

Vilkår for vår anbefaling

Vår anbefaling forutsetter at du gjennomfører prosjektet i tråd med:

- opplysningene gitt i meldeskjemaet og øvrig dokumentasjon
- vår prosjektvurdering, se side 2
- eventuell korrespondanse med oss

Vi forutsetter at du ikke innhenter sensitive personopplysninger.

Meld fra hvis du gjør vesentlige endringer i prosjektet

Dersom prosjektet endrer seg, kan det være nødvendig å sende inn endringsmelding. På våre nettsider finner du svar på hvilke endringer du må melde, samt endringsskjema.

Opplysninger om prosjektet blir lagt ut på våre nettsider og i Meldingsarkivet

Vi har lagt ut opplysninger om prosjektet på nettsidene våre. Alle våre institusjoner har også tilgang til egne prosjekter i Meldingsarkivet.

Vi tar kontakt om status for behandling av personopplysninger ved prosjektslutt

Ved prosjektslutt 11.12.2018 vil vi ta kontakt for å avklare status for behandlingen av personopplysninger.

Se våre nettsider eller ta kontakt dersom du har spørsmål. Vi ønsker lykke til med prosjektet!

Marianne Høgetveit Myhren

Eva J. B. Payne

Kontaktperson: Eva J. B. Payne tlf: 55 58 27 97 / eva.payne@nsd.no

Vedlegg: Prosjektvurdering

Personvernombudet for forskning

Prosjektvurdering - Kommentar

Prosjektnr: 57613

INFORMATION AND CONSENT

According to your notification form the sample (teachers) will receive written and oral information and will give their consent to participate by submitting their answers. The information for the sample that we have received is somewhat incomplete. We ask that the following information is added:

- the name and contact details for the project leader (before the survey is submitted)
- that participation is voluntary
- that information will be treated confidentially and who will have access to it
- the end date of the project and information that any indirectly identifiable personal data will be anonymised by this date.

We ask that you send revised information for informants to personvernombudet@nsd.no. When the information has been revised in accordance with our comments (and sent to us), you can then get started with the project.

INFORMATION SECURITY

The Data Protection Official presupposes that you will process all data according to Høgskolen i Innlandet's internal guidelines/routines for information security.

DATA PROCESSOR

According to the survey form, Qualtrics will be used to carry out the online survey. Qualtrics will therefore be a data processor for the project. If a data processor agreement does not already exist between Høgskolen i Innlandet and the data processor, then a written agreement about how personal data will be processed must be entered into. For advice on what the data processor agreement should contain, please see: https://www.datatilsynet.no/regelverk-og-skjema/veiledere/databehandleravtale/

PUBLICATION

According to your notification form you intend to publish indirectly identifiable personal data. If this is the case, the Data Protection Official presupposes that you will gain explicit consent from each participant to publish their personal data. This information should be added to the information at the start of the survey so that informants are made aware that in submitting their survey answers they are also consenting to their answers being published.

END OF PROJECT AND ANONYMISATION

The estimated end date of the project is 11.12.2018. According to your notification form/information letter you intend to anonymise the collected data by this date. Making the data anonymous entails processing it in such a way that no individuals can be identified. This is done by:

- deleting all direct personal data (such as names/lists of reference numbers)
- deleting/rewriting indirectly identifiable personal data (i.e. an identifying combination of background

variables, such as residence/work place, age and gender)

We draw your attention to the fact that any data processor must also delete personal data linked to the project in its systems. This includes transcriptions, files, logs, links between IP/email addresses and answers.

Sikt

Notification form / 57613. Can corpora be useful in English language teaching in N... / Assessment

Assessment of processing of personal data

Reference number
508565

Assessment type
Standard

Date
02.12.2021

Project title
57613. Can corpora be useful in English language teaching in Norway?

Data controller (institution responsible for the project)
Høgskolen i Innlandet / Fakultet for lærerutdanning og pedagogikk / Institutt for humanistiske fag

Project leader
BARRY KAVANAGH

Project period
01.08.2017 - 31.12.2022

Categories of personal data
General

Legal basis
Consent (General Data Protection Regulation art. 6 nr. 1 a)

The processing of personal data is lawful, so long as it is carried out as stated in the notification form. The legal basis is valid until 31.12.2022.

Notification Form

Comment

NSD has assessed the change registered on 20.11.2021.

We find that the processing of personal data in this project will comply with data protection legislation, so long as it is carried out in accordance with what is documented in the Notification Form and attachments, dated 02.11.2021, as well as in correspondence with NSD. Everything is in place for the processing to continue.

The date for project end is now 31.12.2022.

FOLLOW-UP OF THE PROJECT
NSD will follow-up the project at the planned end date in order to determine whether the processing of personal data has been concluded.

Contact person at NSD: Anne Lene L. Nymoen
Good luck with the project!

Appendix 2. Lawful consent text, and information letters with consent forms

Lawful consent text

English: This is a questionnaire for teachers who teach English in schools in Norway. Answering this 5-minute questionnaire will help us with computer-based tools useful for teachers! This can help teachers and pupils how explore how English is used in everyday life. Participation is voluntary and you do not have to answer all the questions. The information is confidential and only the researcher has access to it. Date will be anonymized in publication. Collected data will be anonymized at the end of the project (31 July 2020). Researcher responsible: Barry Kavanagh 62517238 barry.kavanagh@inn.no

Norsk: Dette er et spørreskjema for lærere som underviser i Engelsk på skolene i Norge. Ved å bevare dette spørreskjemaet som tar ca. 5 minutter, hjelper du oss med et dataverktøy som er nyttig for lærere! Det kan hjelpe lærere og elever med å utforske hvordan englesk brukes til daglig. Deltakelse er frivilling og du er ikke nødt til å svare på alle spørsmålene. Informasjonen er konfidensiell og kun forskeren har tilgang til den. Data blir anonymisert i publisering. Datamaterialet anonymiseres innen prosjekslutt (31. juli 2020). Ansvarlig forsker: Barry Kavanagh 62517238 barry.kavanagh@inn.no

Information letter 1. For follow-up interviews with questionnaire informants

Dear English teacher,

Are you interested in taking part in the research project?

Title: 'Can corpora be useful in English language teaching in Norway?'

This is about participation in a research project where the main purpose is to find out whether corpus linguistics is useful to English language teaching in Norway. In this letter we will give you information about the purpose of the project and what your participation will involve.

Purpose of the project

This is a doctoral thesis consisting of three articles and a binding article. The three research questions are: How familiar are teachers of English in Norway with corpus linguistics? How can teachers of English in Norway use corpora in teaching? How do some teachers of English in Norway use corpora in teaching?

Who is responsible for the research project?

Inland Norway University of Applied Sciences (Høgskolen i Innlandet) is the institution responsible for the project.

Why are you being asked to participate?

You answered the online questionnaire for English teachers:
https://survey.eu.qualtrics.com/jfe/form/SV_8waWUToXSObfk9v
You volunteered to be interviewed and supplied an email address.

What does participation involve for you?

If you choose to take part in the project, you will be interviewed and there will be an audio recording. The interview will be about the contexts in which you have worked with corpora and what you have used corpora for.

Participation is voluntary

Participation in the project is voluntary. All information about you will be anonymized in reporting. If you choose to participate, you can withdraw your consent at any time without giving a reason. All information about you will then be deleted. There will be no negative consequences for you if you chose not to participate or later decide to withdraw.

Your personal privacy – how we will store and use your personal data

We will only use your personal data for the purpose(s) specified in this information letter. We will process your personal data confidentially and in accordance with data protection legislation (the General Data Protection Regulation and Personal Data Act).

Only the researcher Barry Kavanagh will have access to the personal data. No unauthorized persons are able to access the personal data. The researcher will replace your name and contact details with a code. The list of names, contact details and respective codes will be stored separately from the rest of the collected data.

What will happen to your personal data at the end of the research project?

The project is scheduled to end on 31 July 2020. Collected data will be anonymized at the end of the project.

Your rights

So long as you can be identified in the collected data, you have the right to:

- access the personal data that is being processed about you
- request that your personal data is deleted
- request that incorrect personal data about you is corrected/rectified
- receive a copy of your personal data (data portability), and
- send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

What gives us the right to process your personal data?

We will process your personal data based on your consent.

Based on an agreement with Inland Norway University of Applied Sciences, NSD – The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is in accordance with data protection legislation.

Where can I find out more?

If you have questions about the project, or want to exercise your rights, contact:

- Inland Norway University of Applied Sciences via Barry Kavanagh (barry.kavanagh@inn.no) +47 62 51 72 38.
- Vår lokale kontaktperson for personvern i forskning: Anne Sofie Lofthus, forskningsrådgiver, Høgskolen i Innlandet, anne.lofthus@inn.no, telefon: 61 28 82 77.
- NSD – The Norwegian Centre for Research Data AS, by email (personvernombudet@nsd.no) or by telephone: +47 55 58 21 17.

Yours sincerely,

Barry Kavanagh
Project Leader
(Researcher)

Consent form

I have received and understood information about the project ‘Can corpora be useful in English language teaching in Norway?’ and have been given the opportunity to ask questions. I give consent:

- ☐ to participate in an interview

I give consent for my personal data to be processed until the end date of the project, 31 July 2020.

(Signed by participant, date)

Information letter 2. For participants in the corpus seminars.

Dear English teacher,

Are you interested in taking part in the research project?

Title: 'Can corpora be useful in English language teaching in Norway?'

This is about participation in a research project where the main purpose is to find out whether corpus linguistics is useful to English language teaching in Norway. In this letter we will give you information about the purpose of the project and what your participation will involve.

Purpose of the project

This is a doctoral thesis consisting of three articles and a binding article. The three research questions are: How familiar are teachers of English in Norway with corpus linguistics? How can teachers of English in Norway use corpora in teaching? How do some teachers of English in Norway use corpora in teaching?

Who is responsible for the research project?

Inland Norway University of Applied Sciences (Høgskolen i Innlandet) is the institution responsible for the project.

Why are you being asked to participate?

You are being taught the use of language corpora in the course [*course name redacted*].

What does participation involve for you?

If you choose to take part in the project, you will be interviewed and there will be an audio recording. The interview will be about what you think of language corpora. The researcher will also ask to gather data from your worksheets and/or assignments related to corpora.

You can participate in an interview without allowing data to be gathered from the worksheets and/or assignments. You can allow data to be gathered from the worksheets and/or assignments without participating in an interview.

Participation is voluntary

Participation in the project is voluntary. All information about you will be anonymized in reporting. If you choose to participate, you can withdraw your consent at any time without giving a reason. All information about you will then be deleted. There will be no negative consequences for you if you chose not to participate or later decide to withdraw.

Your personal privacy – how we will store and use your personal data

We will only use your personal data for the purpose(s) specified in this information letter. We will process your personal data confidentially and in accordance with data protection legislation (the General Data Protection Regulation and Personal Data Act).

Only the researcher Barry Kavanagh will have access to the personal data. No unauthorized persons are able to access the personal data. The researcher will replace your name and contact details with a code. The list of names, contact details and respective codes will be stored separately from the rest of the collected data.

What will happen to your personal data at the end of the research project?

The project is scheduled to end on 31 July 2020. Collected data will be anonymized at the end of the project.

Your rights

So long as you can be identified in the collected data, you have the right to:

- access the personal data that is being processed about you
- request that your personal data is deleted
- request that incorrect personal data about you is corrected/rectified
- receive a copy of your personal data (data portability), and
- send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

What gives us the right to process your personal data?

We will process your personal data based on your consent.

Based on an agreement with Inland Norway University of Applied Sciences, NSD – The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is in accordance with data protection legislation.

Where can I find out more?

If you have questions about the project, or want to exercise your rights, contact:

- Inland Norway University of Applied Sciences via Barry Kavanagh (barry.kavanagh@inn.no) +47 62 51 72 38.
- Vår lokale kontaktperson for personvern i forskning: Anne Sofie Lofthus, forskningsrådgiver, Høgskolen i Innlandet, anne.lofthus@inn.no, telefon: 61 28 82 77.
- NSD – The Norwegian Centre for Research Data AS, by email (personvernombudet@nsd.no) or by telephone: +47 55 58 21 17.

Yours sincerely,

Barry Kavanagh
Project Leader
(Researcher)

Consent form

I have received and understood information about the project ‘Can corpora be useful in English language teaching in Norway?’ and have been given the opportunity to ask questions. I give consent:

- ☐ to participate in an interview
- ☐ to allow data to be gathered from my language corpora worksheets
- ☐ to allow data to be gathered from my language corpora assignments

I give consent for my personal data to be processed until the end date of the project, 31 July 2020.

(Signed by participant, date)

Information letter 3. For follow-up interviews for seminar participants.

Dear English teacher,

Are you interested in taking part in the research project?

Title: 'Can corpora be useful in English language teaching in Norway?'

This is about participation in a research project where the main purpose is to find out whether corpus linguistics is useful to English language teaching in Norway. In this letter we will give you information about the purpose of the project and what your participation will involve.

Purpose of the project

This is a doctoral thesis consisting of three articles and a binding article. The three research questions are: How familiar are teachers of English in Norway with corpus linguistics? How can teachers of English in Norway use corpora in teaching? How do some teachers of English in Norway use corpora in teaching?

Who is responsible for the research project?

Inland Norway University of Applied Sciences (Høgskolen i Innlandet) is the institution responsible for the project.

Why are you being asked to participate?

You have been taught the use of language corpora in the course [*course name redacted*] and you volunteered for the researcher to conduct follow-up.

What does participation involve for you?

If you choose to take part in the project, you will be interviewed and there will be an audio recording. The interview will be about what you think of language corpora, and how you use them.

Participation is voluntary

Participation in the project is voluntary. All information about you will be anonymized in reporting. If you choose to participate, you can withdraw your consent at any time without giving a reason. All information about you will then be deleted. There will be no negative consequences for you if you chose not to participate or later decide to withdraw.

Your personal privacy – how we will store and use your personal data

We will only use your personal data for the purpose(s) specified in this information letter. We will process your personal data confidentially and in accordance with data protection legislation (the General Data Protection Regulation and Personal Data Act).

Only the researcher Barry Kavanagh will have access to the personal data. No unauthorized persons are able to access the personal data. The researcher will replace your name and contact details with a code. The list of names, contact details and respective codes will be stored separately from the rest of the collected data.

What will happen to your personal data at the end of the research project?

The project is scheduled to end on 31 July 2020. Collected data will be anonymized at the end of the project.

Your rights

So long as you can be identified in the collected data, you have the right to:

- access the personal data that is being processed about you
- request that your personal data is deleted
- request that incorrect personal data about you is corrected/rectified
- receive a copy of your personal data (data portability), and
- send a complaint to the Data Protection Officer or The Norwegian Data Protection Authority regarding the processing of your personal data

What gives us the right to process your personal data?

We will process your personal data based on your consent.

Based on an agreement with Inland Norway University of Applied Sciences, NSD – The Norwegian Centre for Research Data AS has assessed that the processing of personal data in this project is in accordance with data protection legislation.

Where can I find out more?

If you have questions about the project, or want to exercise your rights, contact:

- Inland Norway University of Applied Sciences via Barry Kavanagh (barry.kavanagh@inn.no) +47 62 51 72 38.
- Vår lokale kontaktperson for personvern i forskning: Anne Sofie Lofthus, forskningsrådgiver, Høgskolen i Innlandet, anne.lofthus@inn.no, telefon: 61 28 82 77.
- NSD – The Norwegian Centre for Research Data AS, by email (personvernombudet@nsd.no) or by telephone: +47 55 58 21 17.

Yours sincerely,

Barry Kavanagh
Project Leader
(Researcher)

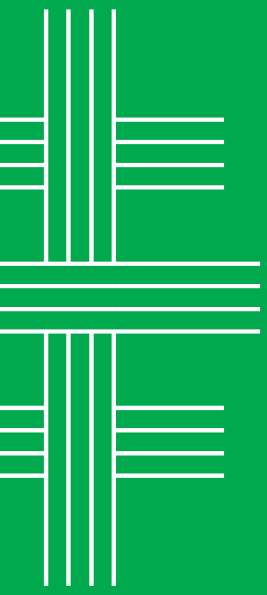
Consent form

I have received and understood information about the project 'Can corpora be useful in English language teaching in Norway?' and have been given the opportunity to ask questions. I give consent:

☐ to participate in an interview

I give consent for my personal data to be processed until the end date of the project, 31 July 2020.

(Signed by participant, date)



Inland Norway
University of
Applied Sciences

This doctoral dissertation focuses on direct pedagogical corpus applications (teachers and/or pupils interacting with corpora), and how these can be useful. The research was conducted in three phases. The first is concerned with how corpora are used by in-service English teachers in Norwegian schools. It comprises a survey of English teachers in primary and secondary school, to discover how widespread the use of corpora is, and follow-up interviews, to obtain the perspectives of corpus-using teachers. The second phase is concerned with what in-service English teachers in Norway find useful about corpora and what they find challenging. It consists of interview data from four teachers who were introduced to corpora through a language course for in-service teachers. The third phase is concerned with how corpora can be used in the English subject in lower secondary school with the current curriculum in Norway. It is composed of corpus exercises, designed by the author, adapted from relevant textbook exercises, and influenced by teacher perspectives from the previous phases.

The dissertation concludes by suggesting a collection of bespoke corpus exercises matched to the curriculum and pupil level, which avoids challenging software or interfaces, uses free and accessible corpora, does not give the impression that the approach is only for linguists, and does not require prior teacher training to use it. Such corpus exercises are to provide teachers with solutions for what is currently required in their English teaching in Norway: authentic language data, explicit language learning, language awareness, critical thinking, in-depth learning, and digital skills.