Inland Norway
University of
Applied Sciences

Faculty of Education

Petter Hagen Karlsen

PhD Dissertation

Teaching and Learning English through Corpus-based Approaches in Norwegian Secondary Schools:

Identifying Obstacles and a Way Forward

PhD Dissertation in Teaching and Teacher Education
2021

Ph.d.-avhandlinger i profesjonsrettede lærerutdanningsfag
PhD Dissertation in Teaching and Teacher Education

Nr. 1 **Irene Trønnes Strøm** – «Jeg er ikke norsk, vet du, jeg er internasjonal».
En etnografisk studie av musikalsk aktørskap blant ungdom i krysskulturelle kontekster

Nr. 2 **James Coburn** – The professional development of English language teachers.
Investigating the design and impact of a national in-service EFL teacher education course

Nr. 3 **Stian Vestby** – Folkelige og distingverte fellesskap.
Gentrifisering av countrykultur i Norge – en festivalstudie

Nr. 4 **Ann Margareth Gustavsen** – Kjønnsforskjeller i sosiale og skolefaglige prestasjoner

Nr. 5 **Camilla Kvaal** – Kryssende musikkopplevelser.
En undersøkelse av samspill i en interkulturell praksis

Nr. 6 **Emmarentia Kirchner** – Motivating and Engaging Readers.
A study of pre-adolescent Namibian primary school readers

Nr. 7 **Gro Løken** – Utsatte grupper og kjønnsforskjeller i spesialundervisning og ordinær opplæring

Nr. 8 **Elin Stengrundet** – Opprørets variasjoner.
Autoritetstematikk i fire dikt av Henrik Wergeland

Nr. 9 **Even Igland Diesen** – «Når vi svinger inn på Macern gjør vi det på ordentlig!»:
En studie i norsk raplyrikk

Nr. 10 **Ellen Nesset Mælan** – Læreres praksis og elevers psykiske helse:
En eksplorerende studie i ungdomsskolen

Nr. 11 **Marit Elise Lyngstad** – English teachers' choices and beliefs about literature in the Norwegian upper secondary classroom

Nr. 12 **Linda Røset** – Physical education and mental health: A study of Norwegian 15-year-olds

Nr. 13 **Anne Randi Fagerlid Festøy** – Vilkår for læring for elevar på 1.-4. steg med vedtak om spesialundervisning

Nr. 14 **Inger-Kristin Larsen Vie** – Nyere, norske biografier for barn og unge:
En studie av biografisjangeren og barn og unges møter med tre biografier

Nr. 15 **Siri Wieberg Klausen –** Fra kritt til programmering: En kritisk diskursanalyse av begrepet digitale ferdigheter i norsk utdanningspolitikk og i norsk videregående opplæring

Nr. 16 **Tove Sommervold** – Litterære lesemåter i ungdomsskolens litteraturundervisning.
En studie av møter mellom ungdomslesere og bildebøker

Nr. 17 **Jonas Yassin Iversen** – Pre-service teachers' first encounter with multilingualism in field placement

Nr. 18 **Lars Bjørke** – The messiness and complexity of pedagogical change.
Teachers' and students' experiences from implementing cooperative learning in physical education

Nr. 19 **Marianne Eek** – Språkhjelpere, investering og sosial rettferdighet.
En undersøkelse av bruk av språkhjelpere i norskopplæringen for voksne nyankomne med liten skolebakgrunn

Nr. 20 **Morten Bjørnebye –** Young children's grounding of mathematical thinking in sensory-motor experiences

Nr. 21 **Petter Hagen Karlsen** – Teaching and Learning English through Corpus-based Approaches in Norwegian Secondary Schools: Identifying Obstacles and a Way Forward

Petter Hagen Karlsen

Teaching and Learning English through Corpus-based Approaches in Norwegian Secondary Schools:

Identifying Obstacles and a Way Forward

PhD Dissertation

2021

Faculty of Education and Pedagogy

Inland Norway University of Applied Sciences

Printed by: Flisa Trykkeri A/S

Place of publication: Elverum

PhD Dissertation in Teaching and Teacher Education no. 21

ISBN trykt: 978-82-8380-301-3

ISBN online: 978-82-8380-302-0

ISSN trykt utgave: 2464-4390

ISSN digital utgave: 2464-4404

Abstract

This doctoral dissertation presents the use of corpus-based approaches to English language learning in upper secondary school in Norway. The research was conducted in two distinct phases. The first phase investigated the pedagogic corpus work of four corpus-trained, in-service teachers and their students' corpus literacy alongside factors that might have influenced this work. Data were collected through a questionnaire to students and teacher interviews. The second phase featured a teacher-researcher collaboration with one of the four aforementioned teachers and two of his first-year, upper-secondary classes to design and implement a corpus-based approach over a two-week period. Data were collected through a case-study design with classroom observations, and subsequent student group interviews.

Previous studies have shown that corpus-based approaches to language learning result in positive learning outcomes; however, most studies are at the tertiary level and designed and conducted by corpus scholars. Meanwhile, data-driven learning in secondary school is "relatively uncharted territory" (Wicher, 2020) and there has been a call for more qualitative studies. The current dissertation sought to contribute knowledge of data-driven learning in the secondary-school context and insight into the processes and opinions of teachers and learners related to pedagogic corpus use.

In the first phase, it was found that the teachers, despite their formal corpus training, had avoided corpus-based approaches in their practice, and few of their students knew anything about corpora. Factors such as teachers' beliefs about their students' digital and linguistic competence and about corpora, teachers' topic focus and epistemic beliefs, and the inaccessibility and cost of corpus applications contributed to their reluctance to introducing their students to corpora. In the second phase, several opportunities for learning were found including instances of metatalk to describe corpus data, peer scaffolding where students helped each other to learn the tool, and teacher scaffolding where the teacher confronted the students with their socio-economic prejudices that arose while working with corpus data from Irish English speakers. However, students' impressions of the tool and process were negatively skewed. Their critique focused on the absence of the teacher, the complexity and aesthetics of the corpus tool's interface and data, and the tool's irrelevance to their learning process.

In addition to the empirical contributions described above, it is argued in the dissertation that there are two major obstacles to data-driven learning that need to be addressed in order for its

application to be normalized in the classroom. These obstacles concern a) the novelty of the approach and the training and mediation required to overcome this novelty, and b) the relevance of the approach to teachers, students, and the curriculum. Inquiry-based education was brought in as a theoretical framework that has considerable overlap with the concepts of data-driven learning but includes a more pronounced social dimension that foster teacher and peer mediation, collaborative learning, and knowledge sharing.

Keywords

Corpus-based approaches, data-driven learning, pedagogic corpora, corpus literacy, Norwegian education, English as a second/foreign language, English language didactics, upper secondary school, pre-tertiary education, subject English in Norway, curriculum renewal, language learning, inquiry-based learning, inquiry.

Sammendrag

Denne doktorgradsavhandlingen tar for seg en studie av korpusbaserte tilnærminger til engelsk språklæring i norsk videregående skole. Forskningen foregikk i to separate faser. I den første fasen ble den pedagogiske korpusbruken til fire videregåendelærere undersøkt, samt deres elevers korpuskjennskap. I tillegg ble faktorer som kan ha påvirket korpusbruken i disse klassene undersøkt. Dataen ble samlet via et spørreskjema til elevene og lærerintervjuer. Den andre fasen innebar et lærer-forsker-samarbeid med en av de ovennevnte lærerne og to av hans første års videregåendeklasser, for å designe og gjennomføre en korpusbasert tilnærming over en to-ukers periode. Dataen ble samlet gjennom en kasusstudie med klasseromsobservasjoner og påfølgende gruppeintervjuer med elevene.

Tidligere forskning viser at korpusbaserte tilnærminger til språklæring har resultert i positivt læringsutbytte, men de fleste studiene er gjort innen høyere utdanning og er designet og gjennomført av korpusforskere. Datadrevet læring i videregående er derimot relativt lite utforsket (Wicher, 2020) og flere kvalitative studier etterspørres av flere forskere. Denne avhandlingen bidrar til mer kunnskap om datadrevet læring i videregåendekonteksten og gir innsikt i prosessene knyttet til pedagogisk bruk av korpus og meningene lærere og elever har i denne sammenhengen.

I den første fasen fant jeg at lærere, til tross for deres formelle korpusutdanning, hadde unngått korpusbaserte tilnærminger i sin egen praksis, og få blant elevene deres visste noe som helst om korpus. Faktorer som læreres oppfatninger [teacher's beliefs] om elevenes digitale og språklige ferdigheter og korpusferdigheter, lærernes temafokus og epistemiske holdninger, og problemer relatert til tilgjengelighet og kostnader, bidro til lærernes motvilje mot å introdusere korpus til elevene. I den andre fasen ble flere språklæringsmuligheter observert, inkludert tilfeller av bruk av metaspråk for å beskrive korpusdataene, elev-scaffolding hvor elevene hjalp hverandre med å forstå verktøyet, og lærer-scaffolding ved at læreren konfronterte elevene med deres sosioøkonomiske fordommer som kom frem mens de arbeidet med korpusdata fra irsk-engelske språkbrukere. Elevene uttrykte likevel negative oppfatninger av korpusverktøyet og undervisningen. Elevenes kritikk omhandlet opplevelsen av læreren som fraværende, kompleksiteten og estetikken til korpusverktøyet og dataene i korpuset, samt verktøyets manglende relevans for læringsprosessen deres.

I tillegg til de ovenfornevnte empiriske bidragene argumenteres det i avhandlingen for at datadrevet læring innebærer to store utfordringer som må løses for at tilnærmingen skal bli

normalisert i klasserommet. Disse utfordringene omhandler nyhetsproblemet [the novelty gap] og den treningen og medieringen som kreves for å løse det, samt korpustilnærmingens relevans for lærere, elever og læreplanen. Utforskende arbeidsmetoder [inquiry-based education] er foreslått som et teoretisk rammeverk som i stor grad overlapper med datadrevet læring, men som også inkluderer en tydeligere sosial dimensjon som innebærer lærer- og elev-mediering, samarbeidende læringsformer, og kunnskapsdeling.

Nøkkelord:

Korpusbaserte tilnærminger, datadrevet læring, pedagogisk korpus, korpus-literacy, norsk utdanning, engelsk som fremmedspråk, engelsk språkdidaktikk, videregående skole, grunnutdanning, engelskfaget i Norge, fagfornyelsen, språklæring, utforskende arbeidsmetoder.

Preface

I began this dissertation back in 2018 and it has been a privilege to be allowed to undertake such a journey. The project has taken me to new places and afforded me valuable experiences that I will remember fondly. There were bumps in the road, but as with any adventure, these challenges tend to fade in hindsight. Along the way, a host of people have contributed to the project in a myriad of ways who deserve recognition and praise.

First and foremost, I am very grateful to the teachers and students who participated in my studies and made the dissertation possible. The four teachers who participated in the first phase of the research not only committed hours of their time to interviews but facilitated the distribution and execution of the questionnaire. In particular, I would like to mention the teacher who open his classroom to me. I am thankful to him and his students for allowing me to interfere in their work and education, for their candid opinions and direct feedback, and for clearing space in an otherwise packed schedule. Their contribution was invaluable.

Alongside the participants, my supervisors have gone above and beyond throughout the entire project. Professor Sylvi Rørvik and professor Marte Monsen have read, commented, guided, proofread, praised, and critiqued my work in extraordinary detail and at an incredible pace. I am convinced that their professionalism, competence, and personal interest in my project were pivotal in the dissertation's completion and in making the process mostly painless. In addition, I got great feedback underway by associate professor Hildegunn Dirdal at the University of Oslo, who provided a structured and comprehensive 50% evaluation, and associate professor Ulrikke Elisabeth Rindal at the University of Oslo, who provided an in-depth and dialogic 90% evaluation of my project. Thank you both for asking good questions, pointing me in the right direction, and giving me new insights and words of encouragement. I am also grateful to associate professor Fiona Farr at the University of Limerick for welcoming me to Ireland, reading and commenting on my second article, inviting me to speak at UL, and for giving me the opportunity to hone my academic authorship skills.

Thank you to my fellow PhD student colleagues for many interesting conversations and social gatherings. A very special thank you to Agnieszka Moraczewska, Eirik Skjelstad, and Synnøve Markeng for the company, support, laughter, and provocative conversations during the final phases of my project.

And finally, I am ever grateful to my friends and family who remained patient, curious, and supportive from start to finish, and who helped distract me in the more frustrating periods of the writing process. Rolf, Else-Kristin, Marie, Magni and Joachim – Thank you!

"You can't get wise with sleep still in your eyes no matter what your dream might be." – Neil Peart

Hamar, August 2021

Petter Hagen Karlsen

Table of contents

List of Tables & Figures

List of Articles

Article 1

Karlsen, P. H., & Monsen, M. (2020). Corpus Literacy and Applications in Norwegian Upper Secondary Schools: Teacher and Learner Perspectives. *Nordic Journal of English Studies: NJES, 19*(1), 118-148.

Article 2

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

Article 3

Karlsen, P. H. (2021). Educational Roles in Corpus-Based Education: From Shift to Diversification. *Nordic Journal of Language Teaching and Learning 9*(1), 1-12. DOI: https://doi.org/10.46364/njltl.v9i1.833

1. Introduction

1.1 Background

This dissertation deals with the use of linguistic corpora (singular: 'corpus') as an educational resource in subject English in secondary school in Norway. The topic was chosen because of the potential impact a corpus-based approach can have on education, an impact of which there is already evidence in higher education (see Boulton & Cobb, 2017), but that has not been observed to the same degree below the tertiary level (see Section 1.3). In fact, the use of corpora for educational purposes in secondary school remains "[…] relatively uncharted territory" (Wicher, 2020, p. 31) and there has been a call for more qualitative studies of this context (Pérez-Paredes, 2020). Moreover, Chambers (2019) pointed to the prevailing research-practice gap within pedagogic corpus consultation, acknowledging the work already done by corpus linguists, applied linguistics researchers, teacher-researchers, and teacher educators, but pointed out that the teachers and learners themselves have the greatest potential to impact the development of corpus-based approaches (p. 471). This dissertation addresses these identified gaps by investigating the use of corpora in the secondary classroom through a teacher-researcher collaboration, and insights from teachers and students. Although these gaps were primarily tackled through empirical data by way of interviews, observations, and a questionnaire, the experiences gained from the research process lead to a re-assessment of the initial approach of open-ended discovery learning by exploring *inquiry-based education* as a viable theoretical framework for corpus-based approaches to language education (see Chapter 3). In other words, in addition to the empirical research based on the theoretical concepts already outlined in corpus-based-learning literature, a substantial part of the discussion will include the examination of corpus-based approaches as *a mode of inquiry*, and in particular how these approaches may be seen to front a change to the roles and conducts of teachers and students alike. Thus, the research gaps presented above are tackled from both an empirical and a theoretical-hypothetical standpoint.

The context of the dissertation's first article is upper-secondary school as a whole, while the second and third articles, which report on the abovementioned collaborative effort to introduce corpora in the classroom, narrow in on two first-year upper-secondary classes. The first year of upper secondary is the last year of mandatory subject English in Norway (see Chapter 2), which means that the students are likely to have varying motivations and interests. It can also be

considered the ‘next step down’ from tertiary education to which corpus-based approaches may trickle.

In this chapter, an introduction of corpus linguistics and the rationale for applying it in language education are covered in Section 1.2. Section 1.3 provides an overview of the state of the art of corpus-based pedagogical approaches, that is, the current state of research in the field. The dissertation’s research design and research questions are briefly outlined in Section 1.4, before its structure is explained in Section 1.5.

1.2 Linguistic Corpora and Language Study

1.2.1 The Nuts and Bolts of Corpora and their Affordances

The current section aims to introduce readers to the basics of corpora in order to give non-corpus-literate readers an initiation into the world of corpus linguistics and portray the available resources, introduce important concepts of corpus linguistics, and discuss the transition from linguistic corpora to corpora for pedagogy and didactic practice. The following section thus provides a brief description of corpora, corpus linguistics, and corpus design, with the intention of giving readers an overview of central terms, concept, and resources.

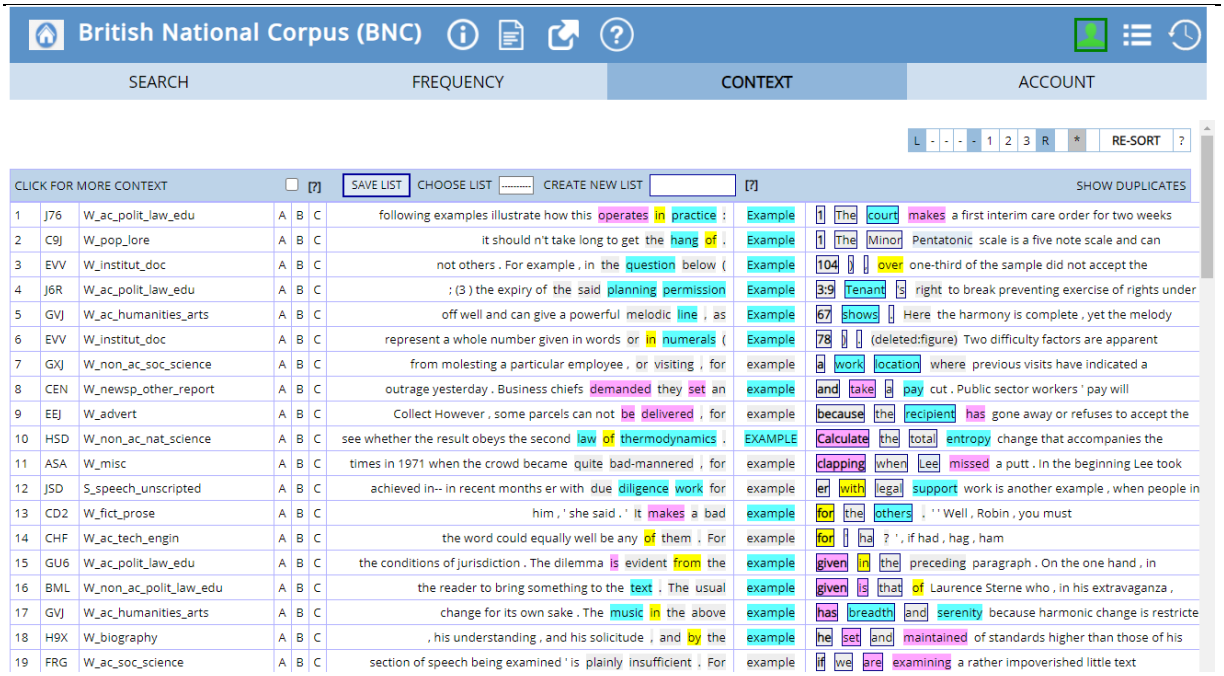
In linguistics, a *corpus* refers to a computer-searchable database of texts, either written or transcribed from speech/video, that is “systematically collected, naturally occurring categories of texts” (Friginal, 2018, p. 11) “sampled to be […] representative of a particular language or language variety” (McEnery, Xiao, & Tono, 2006, p. 5). Corpus linguists seek to describe language systematically through empirical, frequency- and pattern-based approaches (Friginal, 2018, p. 6). The emergence of large, computerized corpora meant that the study of language could move from being based on limited sets of examples, intuition, or introspection, to being based on large datasets containing authentic language in use. In other words, we have moved from “a scarcity to a superfluity” of evidence (Sinclair, 1997, p. 28). Thus, corpora offer unique insights into the nature of language from authentic communication that previously have been difficult or impossible to capture. For instance, a corpus can show the variety in language use across different languages, language variants (sociolects, dialects), genres, registers, age, etc. depending on what types of texts the corpus comprises. Another example is the use of corpora for diachronic studies, where the texts of a corpus represent a language or language variant’s

development over time. Corpus linguistics has also given us ways to examine notoriously difficult language phenomena, such as collocations (words that frequently turn up together in communication), and idiomatic constructions.

Although corpora are only electronic warehouses of text, their usefulness is multiplied through search engines called *Concordancers,* the corpus's annotation scheme, and the metadata provided with the texts. Depending on the application, a concordancer usually provides a range of search options such as KWIC (keyword in context), frequency lists, n-grams, and collocations (Friginal, 2018, p. 46), as well as distribution information. With the KWIC, many examples of your searched word/construction are listed in a central column (see Figure 1) and its immediate context – sometimes referred to as co-text – is displayed on each side. Some concordancers color code or provide tags to the co-text to illustrate the function of the surrounding constructions and help analyze the function of the search term in context. The reading and analysis of concordance searches are referred to as *vertical reading* (Boulton, 2009, p. 40) as you try to find patterns in the list of constructs in multiple contexts. The *frequency list* option shows a tally of the different words in the corpus or parts of the corpus (sub corpus). Frequency information is useful in "[…] both the description of language varieties and in determining what to focus on in a vocabulary lesson" (Friginal, 2018, p. 47). *Collocation* searches supply an overview of which words or phrases tend to appear together in a corpus and how often. *N-grams* involve searches where the target word is shown in the sequence of n number of words that come before and/or after it, which helps one to explore the broader context of word sequences. The degree of searchability of a corpus is linked to its annotation scheme, i.e. the labels put on constructions to aid the concordancers in retrieving them. A common annotation is part-of-speech tagging, where word classes (e.g. nouns, adverbs, prepositions) are labeled; however, some corpora have more complex annotation schemes that increase their affordances. For instance, the ICE-GB [International Corpus of English – Britain component] (ICE-GB, 2020) features a full parsing tag, which means that one can search and view full syntactic trees of sentences within the corpus. Lastly, *metadata* refers to the background information provided with the texts, such as demographic information, genre, register (e.g., spoken or written, academic or colloquial), text elicitation materials (e.g., task wording in the case of student texts), and/or contextual information, etc. In other words, the richer the metadata, the more concrete and detailed are the potential language description of any given study.

Figure 1

Screenshot from a keyword in context (KWIC) search for "example" in the British National Corpus (The British National Corpus, 2007)



1	J76	W_ac_polit_law_edu	A	B	C	following examples illustrate how this operates in practice :	Example	1 The court makes a first interim care order for two weeks
2	C9J	W_pop_lore	A	B	C	it should n't take long to get the hang of .	Example	1 The Minor Pentatonic scale is a five note scale and can
3	EVV	W_institut_doc	A	B	C	not others . For example , in the question below (	Example	104)] over one-third of the sample did not accept the
4	J6R	W_ac_polit_law_edu	A	B	C	; (3) the expiry of the said planning permission	Example	3:9 Tenant s right to break preventing exercise of rights under
5	GVJ	W_ac_humanities_arts	A	B	C	off well and can give a powerful melodic line , as	Example	67 shows . Here the harmony is complete , yet the melody
6	EVV	W_institut_doc	A	B	C	represent a whole number given in words or in numerals (	Example	78)] (deleted:figure) Two difficulty factors are apparent
7	GXJ	W_non_ac_soc_science	A	B	C	from molesting a particular employee , or visiting , for	example	a work location where previous visits have indicated a
8	CEN	W_newsp_other_report	A	B	C	outrage yesterday . Business chiefs demanded they set an	example	and take a pay cut . Public sector workers ' pay will
9	EEJ	W_advert	A	B	C	Collect However , some parcels can not be delivered ; for	example	because the recipient has gone away or refuses to accept the
10	HSD	W_non_ac_nat_science	A	B	C	see whether the result obeys the second law of thermodynamics .	EXAMPLE	Calculate the total entropy change that accompanies the
11	ASA	W_misc	A	B	C	times in 1971 when the crowd became quite bad-mannered , for	example	clapping when Lee missed a putt . In the beginning Lee took
12	JSD	S_speech_unscripted	A	B	C	achieved in-- in recent months er with due diligence work for	example	er with legal support work is another example , when people in
13	CD2	W_fict_prose	A	B	C	him , ' she said . ' It makes a bad	example	for the others , '' Well , Robin , you must
14	CHF	W_ac_tech_engin	A	B	C	the word could equally well be any of them ; For	example	for [ha ? ' , if had , hag , ham
15	GU6	W_ac_polit_law_edu	A	B	C	the conditions of jurisdiction . The dilemma is evident from the	example	given in the preceding paragraph . On the one hand , in
16	BML	W_non_ac_polit_law_edu	A	B	C	the reader to bring something to the text . The usual	example	given is that of Laurence Sterne who , in his extravaganza ,
17	GVJ	W_ac_humanities_arts	A	B	C	change for its own sake . The music in the above	example	has breadth and serenity because harmonic change is restricte
18	H9X	W_biography	A	B	C	, his understanding , and his solicitude , and by the	example	he set and maintained of standards higher than those of his
19	FRG	W_ac_soc_science	A	B	C	section of speech being examined ' is plainly insufficient ; For	example	if we are examining a rather impoverished little text

The different types of corpora may be categorized according to the type(s) of data they contain in order to represent a certain language or language variety. These corpus types have different affordances and have been used to achieve different research and educational goals. *General/reference corpora* are compiled across broad categories and are often the largest sized corpora due to the sheer number of potential texts that may be included in them. Examples of general corpora are the British National Corpus (BNC), which contains 100 million words and seeks to represent British English as a whole across categories such as newspaper articles, fiction, spoken (transcriptions), academic, etc. (The British National Corpus, 2007) and the spoken component of the British National Corpus 2014 (Spoken BNC2014), containing 11.5 million words (Love, Dembry, Hardie, Brezina, & McEnery, 2017). The release of a written component of the BNC2014 is still pending. The Corpus of Contemporary American English is an equivalent for American English with more than one billion words across similar genres (Davies, 2008). A second type of corpus is *specialized corpora*, which contain a specific genre of texts. These can be exclusively texts of business emails from a particular industry or transcriptions of political speeches. One example of a specialized corpus is the Michigan Corpus of Academic Spoken English (MICASE) (Simpson, Briggs, Ovens, & Swales, 2002),

which contains 1.8 million words of transcribed speech from academic contexts. As a third type, *learner corpora* represent the language production of language learners and afford a unique opportunity for researchers to examine the language development, errors, and acquisition of learners. A comprehensive list of learner corpora is provided by the Centre of English Corpus Linguistics at UCLouvain's homepage (Centre for English Corpus Lingustics, 2019). A fourth type is do-it-yourself (DIY) corpora. These are not defined by the language variety they represent per se but are instead corpora assembled by the user to analyze whatever texts s/he wishes to examine. These can be student texts made into a corpus by a teacher who wants to look at typical errors in his students' writing, or a master's student putting together several academic articles from her discipline to make a corpus to aid her in her own writing. These corpora are typically small and personalized. Laurence Anthony has developed a range of popular tools which make DIY corpus creation easy and largely automatized. For instance, TagAnt (Anthony, 2015) is a program that lets you upload and automatically tag self-chosen texts with part-of-speech tags, then, the texts can be loaded into AntConc, (Anthony, 2019), another computer program that allows you to search your texts like a corpus. A fifth type of corpora is *multimedia corpora*. As the name suggests, these corpora present data in more than one medium, for instance as both text and sound files or as both video and text files. Examples of multimedia corpora are the ELISA corpus [English Language Interview Corpus as a Second Language Applications] (Braun, 2006) and the MmCT 1.1 [Multi-modal Corpus Tool 1.1] (Hirata, 2020), which is currently in development. These corpora have been crucial in forming a sixth type, *pedagogic corpora*, which is discussed in Section 1.2.2 below.

Lastly, a final contribution to corpus types is the controversial *web-as-a-corpus*, i.e. searching the Internet as if it were a large corpus. This issue is part of a debate that demands some attention. Kilgarriff and Grefenstette (2003) argue that the web is indeed a corpus in the widest sense, i.e. "a collection of text" (p. 334). They recognize the anarchic nature of the web but point out that although it is only representative of itself, so are corpora (p. 343). The issue of representativeness relates to philosophical questions such as 'what constitutes a language?', and 'what varieties of a given language do one wish to represent?', as well as practical questions such as 'how large a text sample is necessary to represent a (group's) language usage?' or 'how should it be balanced?'. It is beyond the scope of this dissertation to tackle this debate; however, it should be noted that a corpus does offer some control over the research object. While the web is a wild west frontier where anyone can contribute, list, copy and redistribute anything, a corpus is principled in its collection and can therefore be used alongside metadata to clarify the

object one attempts to study (e.g., demographic information). This principled collection strategy allows for the examination of language with clearly provided metadata such as speaker/writer demographics or text-elicitation tools. Moreover, the algorithms that underlie the most frequented search engines (e.g., Google.com) adapt the results to fit your profile and might therefore give a skewed search result. According to Boulton (2015), data-driven learning (see below) does not necessarily require a corpus; rather, it denotes an approach in which *data* are primary. He further notes, "[…] Google + web provides a means to explore huge collections of language data" (Boulton, 2015, p. 268). Nevertheless, Boulton (2015) lists some issues with the web-as-corpus, such as representativeness, the unknown nature of its size and composition, the fact that the web fluctuates, which makes search replicability questionable, the lack of tagging and lemmatizing, and the fact that the web is incredibly noisy – i.e., it contains spam, replications, lists, etc. (p. 273). The web is a powerful tool and an undeniable resource to students. Nevertheless, a corpus can provide a more structured, predictable, and cleaner alternative to represent linguistic or pedagogic phenomena through detailed annotations and rigorous design. It is not a replacement for the web, but an addition to the educational process alongside it.

It is important to point out that "[a corpus] enables the learner/student to explore, to investigate, to generalize, to test hypotheses; but it does not itself initiate or direct the path of learning" (Leech, 1997, p. 5). Corpora have many 'affordances' (Leńko-Szymańska & Boulton, 2015), but they are not inherently pedagogic. Leńko-Szymańska and Boulton (2015) use the example of the affordances of a book as an example: "The realization of these affordances depends on the person's ability to read or his or her ingenuity to use the book to support a wobbling table, etc." (p. 2). In other words, corpora are rich resources with many opportunities tied to them, but it is up to the teachers, students, and researchers to conceptualize, develop and implement them in pedagogically reasonable and creative ways. These observations form the rationale for exploring the potential connections between corpus use and established pedagogic and language-learning theories, as these theories provide an understanding of the educational process and subsequently the feasibility and viability of a corpus-influenced education.

Already, corpora have made an impact on education through material development, curriculum development, and in reference works. These applications of corpora are often hidden to the end user, as they may only utilize data extracted from a corpus or have the corpus as an underlying engine for language queries driven by other websites without the user being aware of the origins of the information. Teachers and students are therefore likely to have encountered corpora in

different forms unknowingly. These types of incidental interactions are known as *indirect applications of corpora* (Leech, 1997; Römer, 2011). However, this dissertation focuses on situations when teachers and/or learners interact with the corpora themselves or *direct applications of corpora* (Leech, 1997; Römer, 2011). The direct application of corpora has become known as *data-driven learning* [DDL], an idea first pioneered by (Johns, 1991). This approach has come to be associated with the use of corpora for pedagogical purposes, often through an inductive learning style (see Chapter 3). One caveat to the 'direct' nature of DDL is the use of printouts containing corpus material, where students work with either raw or engineered data extracted from a corpus instead of accessing the corpus database directly through a computer. This variant is known as a *hands-off* approach, in contrast to the *hands-on* approach where students search the corpus on the computer (Boulton, 2010). In addition, DDL can be plotted on a continuum ranging from teacher-directed to student-led but is fundamentally learner-centered (McEnery & Xiao, 2010, p. 370), and can range from deductive to inductive (Liu & Lei, 2017), but is usually associated with an inductive approach (Flowerdew, 2015, p. 31). Thus, one can imagine an idealized corpus-based education that is student-led and inductive; however, one should take care not to adopt a reductionist view and disregard the multiple affordances of corpora in education.

1.2.2 The Transition into Pedagogic Corpora

As Section 1.2.1 shows, there is a plethora of corpus resources available across spectra of sizes, text types, and affordances. Nevertheless, the uptake of corpora in language classrooms, particularly in pre-tertiary education, has been slow. According to Braun (2007), one of the reasons for the relative absence of corpora in secondary school is the fact that most corpora are made with linguistic research in mind and not for pedagogical purposes. Pérez-Paredes (2020) examined this problem and suggested a distinction between the *possibility scenario* and the *feasibility scenario* (p. 70). The former scenario entails the use of established corpora with students, i.e., adapting corpora meant for linguistics in order to use them in the classroom (ibid.). This approach exploits the wealth of resources that is out there and requires much less time and resources to appropriate. Meanwhile, the latter scenario aims at pedagogic usefulness and is built from the ground up for this purpose. These corpora are called *pedagogic corpora*, which "[…] are topic-driven, they pursue pedagogic rather than linguistic representativeness, and they challenge traditional corpus-search behavior" (Pérez-Paredes, 2020, p. 69). This

approach keeps students' needs front and center through planning, design, and implementation, but has the drawback of requiring an immense process of development while adhering to pedagogic principles. Note that if the feasibility scenario is the only viable option, many resources originally thought to be useful in education may no longer be useful, at least not through direct applications at pre-tertiary levels.

Concordancer design is another consideration related to more learner-friendly corpora. Lee, Lee, and Sert (2015) discussed how four user-interface guidelines could be followed in order to develop an online concordancer that was teacher and learner friendly and intuitive. They build upon Broch's (2009) three already-established guidelines *accessibility*, *simplicity,* and *functionality*, and added to these their own: *manageability* (Lee et al., 2015, pp. 5-9). 'Accessibility' concerns the ease with which the users can access the resource. For Lee et al.'s (2015) specific concordancer, the teacher must have a computer to host the server (p. 5). If the concordancer is integrated into the webpage, however, this requirement can be circumvented. 'Simplicity', to Lee et al. (2015), entails that the interface "is simple enough for users to instantly try the application out themselves" (p. 6) and for teachers to be able to easily upload texts to it (p. 7). This guideline thus pulls into question the intuitiveness of the application. 'Functionality' covers what the tool can do. In the case of Lee et al.'s (2015) application, the primary functions of the tool are to create concordances and mark the target word in red letters for the students and to run several corpus datasets at once for the teacher (p. 8). Arguably, functionality and simplicity potentially influence each other in that more functions (e.g. collocation extraction) and more complexity in functionality (e.g. color-coding nouns blue in the co-text) would reduce simplicity and make the tool less intuitive. Lastly, 'manageability' concerns whether the tool is reliable to use, available, and runs smoothly, for instance by not requiring any plug-ins, only an updated web browser (Lee et al., 2015, p. 9). This guideline appears to be linked to accessibility and how easy the tool is to return to, e.g. how many updates or downloads it requires. In addition to these four guidelines, a fifth one can be added based on the research by Karlsen (in preparation; see Article 2). Karlsen (in preparation) found that secondary-school students commented negatively on the color and design of the interface of a multimedia corpus, as well as on the videos in said corpus. Although it may seem superficial, first impressions can have a powerful impact on students' motivation and investment in the task at hand. Therefore, *aesthetics* is proposed as a fifth guideline for corpus/concordancer interface development that may be more impactful than presumed.

As shown above, the world of corpora is a rich one comprising a variety of resources. There is clearly a continuous effort to adapt and develop tools and materials to both exploit the available assets and to create new ones to accommodate specific needs and particular users. In the following section, the state of the art of research on pedagogic corpus utilization is covered.

1.3 The State of the Art

This section examines the previous empirical research carried out on the use of corpora primarily in the secondary-school context. The studies included in what follows are studies that report on direct applications of corpora (cf. Section 1.2.1), and either focus on or include data on students' and/or teachers' attitudes, perspective, reactions, experience, or beliefs regarding their interactions with corpora for pedagogical purposes. In addition to these users' perspectives, the studies' designs are covered in order to paint a picture of the different approaches taken towards implementing DDL and to show the methods by which these studies typically have collected data.

Table 1 provides an overview of studies in the pre-tertiary context that have the foci outlined above. Certain trends can be observed across several studies. Firstly, the number of investigations of direct applications of corpora in pre-tertiary settings has increased recently, with several studies published in 2020 in connection to Crosthwaite's (2020b) book *Data-Driven Learning for the Next Generation*. A noteworthy exception is Braun's (2007) study, which can be considered a pioneering effort to bring corpora to younger learners (below). An earlier study by Rohrbach (2003) was cited by Braun (2007) as a rare empirical study of a corpus-based approach in the secondary context, but this study could not be retrieved. Secondly, English is the target language for most studies, which means that these are largely English-as-a-foreign/-second-language studies. Thirdly, the methods of data collection are similar; they are largely (quasi-)experimental and have pre- and post-tests, and experimental- and control-group comparisons to measure outcome effects, as well as questionnaires, observations, and interviews to obtain users' perspectives. Additionally, most studies were students' first encounter with corpora (Crosthwaite & Stell, 2020; Di Vito, 2020; Liontou, 2020; Papaioannou, Mattheoudakis, & Agathopoulou, 2020; Szudarski, 2020) that featured both hands-on DDL approaches (Crosthwaite & Stell, 2020), paper-based/hands-off approaches (Di Vito, 2020; Szudarski, 2020), and a mixture of both (Liontou, 2020; Papaioannou et al., 2020). These traits

paint a picture of how research into direct applications of corpora has been conducted in the pre-tertiary setting.

Table 1

Primary characteristics of relevant studies of corpus use in secondary school

Study	N	Age	Location	Target Language	Focus	Data collection methods
Braun (2007)	26	14-16	Germany	English	Lexis/ grammar	Marking papers, computer logs, questionnaire, student comments, observations
Moon and Oh (2018)	285	14	Korea	English	Over-generated *be*	Exp/control*, open-ended survey
Szudarski (2020)	22	16-18	Poland	English	Phrases	Pre-/post-test, exp/control, questionnaire
Papaioannou et al. (2020)	64	15-16	Greece	English	Modal aux	Informal interviews, lesson plan implementation, observations
Di Vito (2020)	127	12-13	Italy	French	Word classes, grammar	Observations and questionnaire
Liontou (2020)	60	12-15	Greece	English	Idioms	Exp/control groups
Crosthwaite and Stell (2020) **	2	10	Australia	English	Revise lexical issues	Observations, screenshots of students' queries

* Exp/control denotes a research design where an experimental and a control group were compared.
**This study was set in primary school.

There are also some divergent features that can be observed across the studies. Learning foci vary in specificity and kind, including both lexis and grammar, error correction and discovery of new patterns, but usually the studies select a narrow language phenomenon as the focus. The variation in participant number between studies is also notable. There is also quite a bit of variety in the corpora employed. Some studies utilized large, pre-existing corpora (e.g., Crosthwaite & Stell, 2020; Liontou, 2020; Papaioannou et al., 2020; Szudarski, 2020), while others made use of corpora created for the particular purpose to fit the students (e.g., Braun, 2007; Di Vito, 2020; Moon & Oh, 2018). The trend seems to lean toward Pérez- Paredes' (2020) possibilities scenario, but with significant contributions to the feasibility scenario as well (cf. Section 1.2.2).

The feedback from students across different studies (see Table 1) that investigated attitudes and reactions to the use of corpora was largely positive. In a case study based on her pedagogic multimedia corpus, Braun (2007) found that the students perceived the corpus-based activities as more useful than the control group's computer exercises and showed interest in and adaptability to the new approach; however, there were issues related to a lack of strategies for interpreting concordances and wordlists, and a prevailing notion among students that they did not learn grammar because no rules were written down. In Di Vito's (2020) study, students described the DDL activities as "new", "relaxing" and "different from the traditional ones" and highlighted increased student interactions and teamwork as positive aspects of the approach (p. 181). Several of her students emphasized clear teacher instructions and peer-to-peer assistance as reasons why the new approach was experienced without much difficulty. Conversely, negative comments were centered around the novelty of the topics tackled in the lessons, the struggle to deduce rules, and one student expressed dislike for the methodology as a whole. Szudarski (2020) found that the students experienced the corpus-based work as useful and that it aided them in avoiding errors, but none of the students expressed that corpus consultation was easy. 44% of students said they would use corpora in the future, but the majority were more reserved. The participating teachers observed obstacles related to students' low language proficiency, students' reliance on traditional approaches (deductive and explicit), and students' expectations of studying language they expected would be necessary in upcoming exams. The students in a study by Papaioannou et al. (2020) preferred corpus examples over those in textbooks, and although first contact with corpora was sometimes awkward, they eventually found querying corpora easy. The students in Moon and Oh's (2018) study were largely positive to the new DDL approach (92%), emphasizing increased grammar consciousness, motivation toward grammar learning, and the authenticity of the native corpus examples as positive aspects. The students also expressed that seeing their own and their peers' writing show up as examples from the learner corpus was interesting and fun. Lower-proficiency learners showed increased persistence with grammar tasks with the DDL approach compared to a more traditional approach. Conversely, negative comments centered on some students' struggle with interpreting basic concordance structures and meanings, and reluctance toward embracing a novel type of instruction. Moon and Oh (2018) argued that essential to their approach's success was the teachers' scaffolding through "personalized help" (p. 59). Lastly, Forti (2020) found that students generally experienced concordances as useful and agreed they could help them avoid future grammar errors. The DDL approach was described as *interesting, useful, energic* and *active* (p. 373). However, there were mixed responses from the students when asked

whether multiple concordance lines were confusing. Some students suggested improvements to the approach such as fewer tests and more peer communication based on their findings from the concordance work.

In a smaller study with two primary school students (age 10), Crosthwaite and Stell (2020) explored the use of corpora in English tutoring using SketchEngine for Language Learners [SKELL] and the British Academic Written English corpus [BAWE]. The tutoring sessions aimed to help students revise lexical issues in their text drafts by using corpora. The students expressed positivity toward the approach, which entailed highlighting errors and querying the corpora. While one student was eager and relatively autonomous from the start, the other student was more reluctant and wanted to observe the tutor first, but eventually became "a self-guided corpus adoptee" (p. 168). According to Crosthwaite and Stell (2020), the tutor was key to the success of the approach. For instance, the tutor showing examples of how to use the wildcard function of the concordancers was enough to make one student learn said function, while the other student required constant reminders from the tutor on how to utilize it, before becoming independent.

A common denominator among studies is the primary focus on the student. The teacher is either featured through informal conversations or the researcher acts as instructor. Feedback from in-service teachers appears to be lacking. Studies examining the state of corpus use among in-service teachers (Callies, 2019; Kavanagh, 2021; Mukherjee, 2006; Vitaz & Poletanovic, 2020) have shown that a corpus-based approach is far from normalized in the classrooms of regular teachers. However, efforts are being made to train pre-service teachers in classroom corpus consultation (e.g., Farr, 2008; Leńko-Szymańska, 2017), which may translate to insights from in-service teachers in the future. On the one hand, our lack of insight into in-service secondary teachers' perceptions of corpora is understandable since few of these practitioners appear to use them. On the other hand, this deficiency is troubling, given the potentially crucial position of the teacher at this level of education.

For comparison, at the tertiary level, Chambers (2007) found shared features across nine qualitative studies such as a tendency to use corpora as a supplementary resource, positive remarks from students that centered on corpus consultation's inductive nature, authenticity, autonomy, self-directed learning style, data relevance, example abundance, and grammar reference possibilities. Meanwhile, negative remarks related to difficulty in using the corpus, the training requirements, frustrations with both too common and too rare words, and descriptions of tasks as tedious, laborious, and time-consuming.

In summary, one can see from the surge of studies in the secondary school setting that there is a growing interest in bringing corpora to younger learners. In addition to valuable empirical research, this interest is further promoted through significant pushes in conceptual developments of corpus-based approaches (e.g., Meunier, 2020; Pérez-Paredes, 2020; Wicher, 2020). The positive reception of corpus-based activities by most learners in these studies, coupled with many conceptual developments, give grounds for optimism. Nevertheless, the dearth of studies and the lack of in-service-teacher focus represent gaps that should not be understated. These contributions and challenges form the backdrop of this dissertation.

1.4 Research Design and Research Questions

1.4.1 Overview of Research Design

The dissertation is article-based and includes three distinct but connected journal articles (Articles 1, 2 & 3) and this introductory chapter. The empirical data upon which each article was grounded were gathered in two distinct phases (see Chapter 4). Phase 1 sought to examine what was already being done in schools by teachers who had had some corpus training from their teacher education, as well as their students' corpus literacy. This phase laid the empirical foundation for the first article. Phase 2 was a case study where corpus resources were introduced into two upper secondary school classes in collaboration with their English teacher, who had also participated in phase 1. The findings of this phase became the basis of the second and third articles. Each article has its own set of research questions that can be seen as components of the overarching research question of the dissertation (see Section 1.4.2). While the first phase examined the selected participants' knowledge and experiences with corpora prior to researcher intervention, the second phase focused on the teacher and students' impressions and conduct during and after the implementation of corpora.

1.4.2 Research Questions

The aims of the dissertation are to ascertain the experiences of students and teachers in their encounters with corpora and corpus-based approaches in subject English in Norway, and to add to the practical and theoretical understanding of the corpus-research community. The dissertation does not seek directly to measure the outcomes or the effectiveness of its

participants' interaction with corpora; instead, the focus is on investigating new ways forward for direct applications of corpora in pre-tertiary education by being cognizant of practitioner voices and classroom realities. The following question is therefore posed to frame the dissertation as a whole:

> ***How can corpus-based approaches be integrated into Norwegian secondary schools and how are they received by the users?***

In order to answer this question, each article addresses different aspects of the research. The following sub-questions were formulated:

- **Article 1**: Corpus Literacy and Applications in Norwegian Upper Secondary Schools: Teacher and Learner Perspectives
    - *How familiar are upper secondary school students with corpora?*
    - *What beliefs do teachers express about corpora as a pedagogical tool?*
- **Article 2:** Integrating Multimedia Corpora in the Secondary School Classroom in Norway
    - *How can pedagogic corpora be applied in an upper secondary school and how is this experienced and approached by the teacher and learners?*
    - *What learning opportunities and challenges emerge when introducing corpora directly in the EFL classroom?*
- **Article 3:** Educational Roles in Corpus-Based Education: from Shift to Diversification
    - *How do DDL proponents' assumptions about the upper secondary classroom and its educational roles align with the experiences and opinions of students?*
    - *How can perspectives from inquiry-based education and student-centered teaching inform the conceptualization to educational roles in DDL?*

Corpus linguistics provided the tools and resources central to this project, as well as theoretical and practical perspectives on the integration of corpus-based approaches in education through DDL. These perspectives, however, were supplemented by didactic concepts such as discovery learning and inquiry-based education (see Chapter 3), which allowed for the placement of corpus-based approaches to education into more general pedagogic frameworks. Furthermore, the connections to SLA grounded the research in contemporary theories of language acquisition.

1.5 Structure of the Introductory Chapter

Following this introduction in Chapter 1, the dissertation is structured as follows: Chapter 2 outlines the Norwegian secondary-school context and recent curricular developments in Norway. Chapter 3 provides an overview of the theoretical perspectives that accompany DDL as an approach to language teaching and learning, and presents the theoretical framework of the dissertation, namely inquiry-based education. In Chapter 4, the methodology of the dissertation is presented, including the data collection methods utilized in each of the research phases, the approach to data analysis, and information about the participants. Chapter 5 gives a brief summary of each article. The connections between the articles and the overall findings of the research conducted for this dissertation are discussed in Chapter 6 alongside an overview of contributions, future directions, and concluding remarks. Articles 1, 2 and 3 are placed toward the end of the dissertation in the chapter Dissertation Articles. Appendices 1-5 are the data-collection materials (i.e., interview guides and questionnaires), lesson plans, and the tasks used in the case study. Appendices 6-7 contain the NSD evaluation and consent forms.

2. The Norwegian Upper-Secondary-School Context

2.1 Chapter Introduction

This chapter covers English as a subject in Norwegian upper-secondary education and frames the dissertation against the backdrop of the English subject curriculum, with a particular focus on language learning, and current curricular developments in Norway. A new curricular reform is currently being implemented that brings changes in both the core curriculum and subject curricula (see Section 2.3.3). In addition, the elevation of teacher training to a master's program from 2017 (outlined in White Paper nr. 11, 2008-2009) signals shifts in the educational landscape of Norway that have a bearing on subject English as well as the teaching profession and the educational system as a whole. The increased academization of teacher training alongside the focus on exploratory approaches, in-depth learning and discipline-specific working methods promoted in the new educational reform require new ways of thinking about teaching and learning that open up for an inquiry- and corpus-based approach like the one suggested in this dissertation (see Chapter 3). The aim of the present chapter is therefore two-fold: (1) it aims to inform the reader of the educational and curricular context in which the research took place, and (2) it aims to show the connections between the dissertation's objective and the objectives of the new curricular developments in Norway (see Section 2.3.4). The *way forward* presented in this dissertation aligns well with these new developments and speaks to its future relevance. It follows that the main focus of the present chapter had to be shared between the operational curricula at the time of data collection (LK06; Section 2.3.2) – as the research design was based on these documents – and the newly-implemented curriculum (LK20; Section 2.3.4).

Following this introduction, Section 2.2 gives a brief overview of upper-secondary education in Norway. Section 2.3 presents the structure, concepts, and reforms of the upper-secondary school curricula focusing on the road to competency-driven curricula and local freedom (Section 2.3.1), the previous curriculum LK06 (Sections 2.3.2) and the newly-implemented curriculum (Sections 2.3.3 and 2.3.4). Section 2.4 summarizes the chapter.

2.2 The Norwegian Student and Upper-Secondary Education in Norway

Upper-secondary school is voluntary in Norway; however, the Norwegian Directorate of Education and Training reported that 93 percent of teenagers aged 16-18 were enrolled in upper-secondary education during the school years 2018-2019 and 2019-2020 (Utdanningsdirektoratet, 2020d). While compulsory education in Norway comprises primary school (year 1-7) and lower-secondary school (year 8-10), these numbers show that further education through upper-secondary school is the norm. Notably, failing to complete upper-secondary school is linked to unemployment, no further education, and salary inequity (Bergsli, 2013, pp. 34-35). In addition, high dropout rates have a huge personal and societal cost (Lillejord et al., 2015). Thus, attending upper-secondary school is the recommended and encouraged path for Norwegian students to take.

Norwegian students may choose between *general studies* programs and *vocational* programs. The former is a three-year education (year 11-13) and includes five 'directions' one can choose between: *Music, dance and drama*; *Sports*; *Art, design and architecture; Media and communication;* and General studies intended to qualify for higher education. In addition, general-studies students select program-specific courses in their second year in which to specialize, which are either a) *the sciences,* b) *language, society, and economy* or direction-specific programs (e.g., further specialization in sports subjects for Sports programs). Vocational study programs are normally two years of schooling followed up by two years of apprenticeship – or the 2+2 model – with education geared toward a specific kind of profession (e.g., *Construction* or *Technology and industry*). During the first year for general-studies students and the two first years for vocational-studies students, English is "a compulsory common core subject" (Brevik, Skarpaas, & Isaksen, 2020, p. 65) alongside Norwegian, mathematics, PE, etc. This means that most Norwegian students today are taught English in schools from year 1 through year 11 (or 12 for vocational students). General-studies students may choose to specialize further in English during the two consecutive school years. Since most of the students who are featured in this dissertation were first-year upper-secondary-school students – a few classes were second-year – they were students who had English as a compulsory subject. Some of these students were planning to choose English the following year, while some were second-year students who had already done so (see Chapter 4 for student meta-data). Due to the fact that the case study of this dissertation involved only first-year upper-

secondary students, the curriculum concerning common-core English will be covered and not the one from further specialization in English.

It is also worth mentioning that Norwegians are generally considered competent English users (Brevik, 2019; Crystal, 2012). For instance, Norway was ranked fifth in English proficiency in 2020 among non-native countries and placed within the 'very high proficiency' category (Education First, 2020). The aforementioned amount of formal English training students receive in school can explain this high proficiency in part. In addition, Norwegians are exposed to a lot of English outside the classroom, or *extramural English* (Brevik, 2019), through online platforms like YouTube and Instagram, and undubbed English media (e.g., movies, shows, and television). In addition, the relatively high socio-economic status of many Norwegians allows for frequent international travel where English is used to communicate. This means that if students choose to attend upper secondary – as most do – they are likely to bring with them a considerable amount of informal language knowledge and language-learning experience from authentic sources into the classroom. The combination of formal training, relatively high language proficiency and a familiarity with engaging with authentic texts could lay the groundwork for the more structured approach to authentic language that corpus-based approaches offer. On the face of it, one might assume that these experiences somewhat alleviate the concern that corpora contain language that is too messy and difficult for the learners, although one should be cautious when making assumptions across all students and contexts. Nevertheless, both teachers and educational scholars should be cognizant of students' potential and experiences from both formal and informal settings when designing, planning, and implementing new classroom approaches. With these observations in mind, we move next to the Norwegian curricula as the foundation of the school's objectives and the guiding document for the current research.

2.3 The Norwegian Curriculum: Content, Structure and Reforms

This section provides an overview over Norwegian curricula. First, a few central historical developments that have shaped the school system of today are covered in order to provide a fuller picture of the context in which the research took place. Second, the content and structure of the current curriculum (LK06) and the transition into the new curriculum (LK20), as well as how they relate to the current project, are discussed in detail. The students who participated in this dissertation's studies would have been exposed to LK06 throughout their education and

LK06 was the operative framework when the dissertation's case study was conducted (see Chapter 4). Since the curriculum outlines the schools' objectives and can be considered the teachers' job description (Munden & Sandhaug, 2017), the case study was designed to help the teacher and his students reach a selection of curriculum aims. Furthermore, the teachers, who were all in their late twenties or early thirties, would have gone through teacher training with LK06 as their curricular framework, worked with said curriculum as in-service teachers, and will have to implement LK20 in their practice in the years to come. As a result, the main focus of the following sections is on LK06 as the operative curriculum at the point of data collection, and on LK20, as it represents the school of the future and frames the continued relevance of this dissertation in the Norwegian context.

2.3.1 Toward Competence-Focus Curricula and Local Freedom

The Norwegian educational system and curricula have undergone several major changes through a series of educational reforms. This section focuses on the road toward the current competency-centered curricula and lays out some key concepts along the way. The competence-focus curriculum and the local freedom it permits form the foundation for the research design of this dissertation. Note that the reforms involved several significant changes to school structure and organization that are beyond the scope of this dissertation to address.

The aims-centered curriculum model was first introduced with Reform 94[1], which meant that *competencies*, and not specific *content*, became the subject curricula's central category (Markussen, 2007). This change entailed a focus on providing students with a broader basic competence that would qualify for a series of specialization at higher levels, such as the ability to acquire and apply knowledge. These aims defined what students were supposed to have achieved through working with the subject and were organized as larger, overarching aims with a longer series of competencies described as subcategories of these aims. The aims-based model alongside the core curriculum of Reform 94, which described the school's value system and objective, carried over into LK06, but new subject curricula were created, and the curricula of primary, lower-secondary, and upper-secondary school were merged into one framework (Hølleland, 2007) while the content guidelines were made vaguer (Isnes, 2007). These changes provided a clearer common thread throughout compulsory education and upper secondary.

[1] The numbers in the reform and curriculum names denote the year they were first put into effect.

They also meant that the teachers should ideally be able to select the content, methods, and resources necessary to reach the aims and ideals set by the government (ibid.), which afforded schools significant methodological and organizational freedom (Hølleland, 2007, p. 19). An ambition with LK06 was to strengthen certain *basic skills* after the first PISA [Programme for International Student Assessment] test results in 2001 revealed unsatisfactory performances by Norwegian students (Fladmoe, 2013; Haugsbakk, 2013). This strengthening was achieved by introducing oral skills, writing, reading, numeracy, and digital skills as *basic skills* not only in the English subject but in every subject (Engelsen & Karseth, 2007; Haugsbakk, 2013; Isnes, 2007). Thus, the competence aims of LK06 were developed in every subject with a view to these basic skills. They were no longer organized under overarching aims, but under *Main subject areas* defined for each particular subject (Utdanningsdirektoratet, 2013; see Sections 2.3.2 and 2.3.3 where the English subject curricula are discussed in more detail). The relevant curricular developments of LK20 are discussed in Section 2.3.3 as they entail considerable changes to both the core and subject curricula, but a few points are touched upon here. The basic skills from LK06 have been retained in LK20, but with *oral skills* renamed to *verbal skills*. The aims-based model also remains, but the competence aims are no longer categorized under main subject areas. Instead, the aims are listed in one concurrent list and the number of aims has been reduced from 27 to 17. On the one hand, these developments arguably lead to even vaguer guidelines for teachers, who are afforded greater professional autonomy and responsibility in selecting content, methods, and activities. On the other hand, the focus on discipline-specific education and student-active approaches in the new curriculum can be interpreted as methodological guidelines, albeit vague ones, that introduce new requirements that potentially restrict teacher autonomy.

This section shows that the Norwegian curriculum aims have become increasingly broad and that local freedom – i.e., the freedom of schools and teachers to determine their organization, contents, and methods – has become more prominent. It is in this space of professional autonomy, which simultaneously affords local opportunity for new approaches and responsibility for effective education, that teachers find themselves. It is argued in this dissertation that a collaborative effort between researcher and teacher can serve to help teachers and students reach their goals while staying cognizant of their different needs, and thus help teachers maneuver this space of professional autonomy. This collaboration also involves a responsibility of educational and corpus scholars to keep the objectives of the school in mind, be critical of one's own approaches, and avoid agenda-driven interventions. Such critical

collaborations can then aid teachers, students, and researchers in discovering what works, and avoid the pitfalls of "anything goes". Next, we turn to the English subject curriculum of LK06, which describes the subject and aims of the students and teachers who participated in this dissertation.

2.3.2 Subject English and Language Learning in LK06

The English subject curriculum of LK06 describes English as "both a tool and a way of gaining knowledge and personal insight" (Utdanningsdirektoratet, 2013, p. 2) and outlines the *purpose* of the English subject, the *main subject areas* (*language learning*, *oral communication*, *written communication,* and *culture, society and literature*), the teaching hours for each stage (140 hours for general studies, and 84 and 56 hours for year one and two for vocational studies respectively), the *basic skills* (*writing, reading, oral skills, digital skills and numeracy*), the *competence aims* organized under the main subject areas, and a brief overview of *summative assessments* (Utdanningsdirektoratet, 2013). This section will primarily focus on the competence aims, the subject area of language learning, and the subject's purpose, as these elements are particularly relevant to the dissertation.

The purpose of English is described in terms of being a "universal language" and "world language", where learning vocabulary and the "systems of the English language" are emphasized alongside "cultural norms and conventions" when using the language for communication (Utdanningsdirektoratet, 2013, p. 2). Munden and Sandhaug (2017) argue that the curriculum promotes two main reasons for learning English, namely, "to communicate, and to learn about others" (p. 50). Notably, with a strong focus on basic skills and communication, the acquisition of language form(s) becomes secondary or a means to a communicative end. For instance, vocabulary learning appears only in two aims: "understand and use a wide general vocabulary and an academic vocabulary related to his/her own educational programme" under *oral communication*, and "understand and use an extensive general vocabulary and an academic vocabulary related to one's education programme" under *written communication* (Utdanningsdirektoratet, 2013, p. 10). Meanwhile, the word *grammar* is not mentioned in the competence aims directly, but one aim reads, "use patterns for orthography, word inflection and varied sentence and text construction to produce texts" under *written communication* (ibid.). It is not given that teachers should focus particularly on patterns of language or teach grammar and vocabulary explicitly. Consequently, the focus on patterns associated with data-driven

learning may not be experienced as directly relevant by English language teachers – this, in turn, represents a barrier for the implementation of corpus-based approaches, as the teachers' aforementioned local freedom can essentially make them methodological gatekeepers (cf. Section 2.3.1).

The curriculum's focus on communication is influenced by Hymes' concept *communicative competence*, which includes *language knowledge* and *the ability for [language] use* and thus encompasses cultural, situational, and pragmatic contexts in addition to vocabulary and grammar acquisition (Rindal, 2014, p. 5). This idea sets *intelligibility* and *appropriateness* as language learning goals as opposed to native-like performance (Rindal, 2020, p. 34). However, the goals of intelligibility and appropriateness may leave one with several questions such as 'intelligible to whom?', 'appropriate in what context?' and methodological issues such as how one is to select language examples and facilitate language exposure to ensure that students become competent communicators. These educational gaps are left up to the teacher to fill and, as Rindal (2020) points out, intelligibility and appropriateness can be interpreted as intelligible and appropriate in the perception of native speakers (p. 34). In other words, devoid of clear guidelines on what language model one should strive for, teachers might default to idealized forms of British English (Received Pronunciation) or American English (General American), which have historically enjoyed privileged positions in the classroom and textbooks, or they might elect to teach on the basis of topics and leave language learning as an implicit acquisitional process – as seemed to be the case with the teachers interviewed in Karlsen and Monsen (2020; Article 1).

If one examines language learning in particular, as it is described in the purpose section of the English subject curriculum for all levels, there is no mention of grammar, vocabulary, or pattern learning:

> Language learning occurs while encountering a diversity of texts, where the concept of text is used in the broadest sense of the word. It involves oral and written representations in different combinations and a range of oral and written texts from digital media. When we are aware of the strategies that are used to learn a language, and strategies that help us to understand and to be understood, the acquisition of knowledge and skills becomes easier and more meaningful. It is also important to establish our own goals for learning, to determine how these can be reached and to assess the way we use the language. Learning English will contribute to multilingualism and can be an important part of our personal development. (Utdanningsdirektoratet, 2013, p. 2)

Meanwhile, for upper-secondary school, there are three competence aims categorized under the main subject area of *language learning*, which say that the student should be able to: (a) "evaluate and use different situations, working methods and learning strategies to further develop one's English-language skills", (b) "evaluate own progress in learning English" and (c) "evaluate different digital resources and other aids critically and independently, and use them in own language learning" (Utdanningsdirektoratet, 2013, p. 10). The focus is on practical and meta-cognitive skills and abilities such as using and evaluating appropriate learning strategies including digital tools. The goals of language learning are to understand and be understood, which resonate with the terms *intelligibility* and *appropriateness*. Arguably, the point is not to teach or learn content, but to equip students with a diverse set of skills and the knowledge to actively partake in their own learning processes in and beyond the classroom.

These foci represent both opportunities and hurdles for the introduction of a corpus-based approach that have to be considered. One hurdle is the aforementioned pattern- and form-focus of data-driven learning. Since corpora are essentially collections of texts and a primary function of corpora is to search for linguistic patterns, it could be too linguistically oriented juxtaposed with the communicatively-oriented curriculum. Another hurdle is that data-driven learning can be individualistic. The examination of corpus data on a computer is not conducive to communication in and of itself, particularly with respect to oral skills, although corpora can be a window to observe other people's authentic communication. There are, however, several other aspects of data-driven learning that do coincide with the curriculum. One opportunity corpora provide is a language-learning strategy that does not focus on particular language content, but on skills students can cultivate for further language inquiry. This aligns well with competence aim (a) above. A second opportunity is that corpora offer access to a diversity of texts, which can be both oral and written, and multimedial – as was the case with the corpora used in this dissertation. A third opportunity afforded by a data-driven learning approach is the access to typically neglected variations of English (see Pérez-Paredes, 2020) and the problematization of idealized standards of English through evidence-based methods. In data-driven learning literature, these benefits are referred to as awareness-raising (e.g., Leńko-Szymańska & Boulton, 2015) and align with competence aim (c) above. Two final opportunities for data-driven learning to fulfill curricular demands are the partial focus on frequencies, and the computerization of corpora, which afford ways of working with the basic skills of numeracy and digital skills. Thus, corpora are gateways to diverse language information that can promote knowledge construction through diverse and culturally-situated language examples. Note that

corpora do not "initiate or direct the path of learning" (Leech, 1997, p. 5; cf. Chapter 3), but represent tools and resources that need to be implemented in manners that are practically, theoretically, and didactically sound (cf. Chapters 3 and 6).

This section has covered the English subject curriculum that was the foundation of the dissertation's research and outlined ways in which its focus and views on language learning are congruent or incongruent with a corpus-based approach. We turn next to the new developments and current educational reform in Norway.

2.3.3 LK20: Curriculum and Subject Renewal

The new curriculum LK20 was based on the official 2015 report to the Ministry of Education and Training, *The School of the Future: Renewal of subjects and competences* (Ludvigsen et al., 2015), written by the Ludvigsen committee, which was appointed by the government in 2013. The aim of the report was "to assess the subjects in primary and secondary education and training in terms of the requirements for competences in future working life and society" (Ludvigsen et al., 2015). As of 2021, the new curriculum has been partially implemented for upper-secondary school. The implementation began in 2020 and the aim is for it to be completed for all levels from primary up to and including upper-secondary school by the end of the 2023 school year (Utdanningsdirektoratet, 2020c). Both the subject curricula and the core curriculum have been renewed; however, the curricula remain competence-centered with the subjects and their competence aims at their core. Moreover, the report advises a continuation of locally determined subject content, working methods, organization, and teachers' professional autonomy. The continued collaboration between teachers, schools and research communities thus remains pertinent and is even reinforced through the new focus on discipline-specific competencies and research-emulated working methods (discussed below). In addition, *Health and life skills, Democracy and citizenship,* and *Sustainable development* are introduced as cross-subject themes that influence the subject curricula to varying degrees (Utdanningsdirektoratet, 2020a). Broadly speaking, the former centers on students' physical and mental health, and their ability to make good life choices, the medial involves learning about the requirements, values, and rules necessary to participate in a democratic society, and the latter is linked to the consequences of human living and consumption, and how social, economic, and environmental situations are connected. Out of the three, *Democracy and citizenship* is featured in the new English subject curriculum where it is realized by cultivating

an understanding in the students that their world view is culture-dependent, and that English can be a tool for interacting with people from diverse backgrounds.

Several of the more prominent developments and foci outlined in the new report are particularly central to this dissertation and speak to its future relevance. These foci are *the four areas of competence* and *in-depth learning* (Ludvigsen et al., 2015). One area of competence is *subject-specific competence*, which relates to learning the concepts, principles, and scientific methods of the particular discipline – in this case *languages* – in order to give students the necessary resources to tackle their future lives. The committee recognizes English as an international language, and linguistic competence as closely related to communication, interaction, and participation (p. 26). Another competence area, *being able to explore and create*, has its foundations in discipline-specific scientific methods, critical thinking, problem solving and "an exploratory approach to knowledge" (p. 33). The committee asserts, "Young people are by nature inquisitive and exploring, but curiosity must be stimulated to be developed" (ibid.). These two competence areas are both related to scientific methods and thinking and are highly congruent with inquiry-based approaches to learning (see Chapter 3). Data-driven learning as a mode of inquiry becomes one way in which (1) subject-specific, scientific ways of working can enter the language classroom, and (2) the cross-curricular ambitions of the new curriculum to promote transferable, research-like skills and different strategies can be worked with in the English classroom. According to the committee, scientific ways of working, critical thinking, and problem-solving go hand-in-hand and include assessing the validity of information and arguments, analyze, judge the relevance of knowledge and methods, test, explore, acknowledge that one may not find answers right away, and work in an investigative manner (Ludvigsen et al., 2015, p. 36). These strategies alongside active student participation, collaboration, and self-assessment are deemed conducive to *in-depth learning* – another key word in the report and the new core curriculum. These elements are found in both data-driven learning and inquiry-based approaches (see Chapter 3). Notably, the report emphasizes that "[…] even if the pupils are practicing at working independently, school and the teachers are still responsible for facilitating the pupils' learning processes" (Ludvigsen et al., 2015, p. 29). This statement touches on the teacher's role and sets a precedence for the facilitator role, which will be discussed in Chapter 6 of the dissertation.

The other two competence areas are more indirectly addressed in the dissertation. *Being able to communicate, interact and participate* encompasses the previous basic skills *reading, writing* and *verbal skills* (previously *oral skills*). This area relates to inquiry in that it is linked to

collaboration and one's ability to communicate, argue, and debate one's points. It also shows that communicative language teaching still acts as a prominent influence on the curriculum. Arguably, the same hurdles and opportunities identified in relation to data-driven learning in Section 2.3.2 are applicable to this curricular framework as well; in particular, the danger of too-individualistic working methods becomes more highlighted with the reinforced importance of interaction and participation in the new curricula. It should be noted that the two final former basic skills *numeracy* and *digital skills* are also emphasized in the new curriculum. Lastly, *being able to learn* as the final competence area involves metacognition, i.e., "reflecting on one's own thinking in different contexts" and self-regulated learning, i.e., "learn to take initiative and control parts of their own learning" (Ludvigsen et al., 2015, pp. 28-29), which shows a socio-constructivist influence in the report. These theoretical constructs have been linked to data-driven learning (cf. Flowerdew, 2015; Chapter 3) and inquiry (see Chapter 3).

The connections between an inquiry-based approach, of which data-driven learning can represent one node in a language context, and the foci of the new curriculum will be addressed further in Chapters 3 and 6. In the following section, we take a closer look at the new English subject curriculum (LK20).

2.3.4 Subject English and Language Learning in LK20

The new curriculum framework of LK20 represents several major changes in both structure and content of the *Curriculum in English*, which was passed November 2019 and implemented from August 1st, 2020 (Utdanningsdirektoratet, 2020b). The *purpose* of the subject is replaced by *Relevance and central values* and *Core elements*. The communicative focus remains, but a strengthened English-as-a-lingua-franca [ELF] influence is found, as students must learn to communicate with others "[…] regardless of cultural or linguistic background" (Utdanningsdirektoratet, 2020b, p. 2). According to Rindal (2020), ELF represents a new impulse to the status of English in Norway that emphasizes the fluidity and hybridity of languages over established native models. ELF is based on the notion that cultural and linguistic practices flow across cultures and draw on multiple and complex resources (De Bartolo, 2020, p. 614). However, it does not provide a coherent model of English and may fall into the same traps as communicative language teaching wherein teachers default to standardized native examples. Corpora comprised of non-native or neglected variants may offer data that make English visible as a lingua franca.

The view of English both for communication and as a tool for learning also remains in the new subject curriculum; however, the new description promotes "[…] an exploratory approach to language, communication patterns, lifestyles, ways of thinking and social conditions […]" (Utdanningsdirektoratet, 2020b, p. 2) as well as student curiosity and engagement. These descriptors reflect the scientific ways of thinking proposed in the committee's report (cf. Section 2.3.2) and are congruent with inquiry-based and research-emulating ideas. The new curriculum gives a brief description of both formative and summative *assessments*. Lastly, the organization of competence aims under main subject areas has been removed and the new competence aims are now presented together, but with one set for vocational and one set for general-studies programs. There are now 17 aims each for general studies and vocational studies, down from 27 aims where general and vocational studies were combined; some of these new aims are identical for both general and vocational studies, while others are similar but more geared toward higher education for the former and work life for the latter.

Language learning alongside *Communication* and *Working with texts in English* are the core elements of subject English in LK20. The subject curriculum describes language learning thus:

> Language learning refers to developing language awareness and knowledge of English as a system, and the ability to use language learning strategies. Learning the pronunciation of phonemes, and learning vocabulary, word structure, syntax and text composition gives the pupils choices and possibilities in their communication and interaction. Language learning refers to identifying connections between English and other languages the pupils know, and to understanding how English is structured. (Utdanningsdirektoratet, 2020b, pp. 2-3).

This strengthened presence of lexical and grammatical terminology reinforces the position of linguistic forms in subject English alongside the prevailing focus on communication and language-learning strategies. Moreover, the core element 'working with texts in English' entails students working with texts "[that] can contain writing, pictures, audio, drawings, graphs, numbers and other forms of expression that are combined to enhance and present the message" (ibid.). The combined effort to emphasize the building blocks of language and multimedia texts appears to present especially fertile soil in which to plant the seeds of corpora and, in particular, multimedia corpora.

Lastly, the renewed competence aims will not be covered to the same extent here as the LK06's aims were, as the latter were featured explicitly in the dissertation's second research phase (cf. Chapter 4). However, a few interesting developments should be noted that are relevant for data-driven learning as a mode of inquiry. One feature is the aforementioned reduction of aims from

27 to 17, which affords the teacher more time to work with each aim and signals the prioritization of in-depth learning. Another feature is the exclusion of any overarching categories into which the aims are placed, which suggests freedom to work with a given aim in relation to different basic skills and core elements. Finally, a quick comparison between the verbiage in the previous and the new curricula shows that certain of the new aims are geared toward exploration and science-inspired methods. The verb usage hints at this point with the appearance of "explore", "reflect", and "compare" for both the vocational and general studies aims in LK20, while the general studies aims also feature "analyze" and "reason". The word "interpret" has been carried over from LK06 to general studies in LK20. These verb additions alongside aims such as "read and compare different factual texts on the same topic from different sources and critically assess the reliability of the sources" and "use different sources in a critical, appropriate and verifiable manner" (Utdanningsdirektoratet, 2020b, pp. 4-5), and the vagueness of their wording exemplify ways in which the competences aimed at in the curriculum are infused with academic and scientific ideas and how the teacher is simultaneously awarded more interpretative freedom and discipline-specific methodological requirements.

2.4. Chapter Summary

This chapter introduced readers to the Norwegian upper secondary school context and concepts and developments of the curricula. The aim was to show the environment in which this dissertation is situated as well as its relevance for the time to come. English in Norway is a fuzzy category but with a central position in society and education. The English subject in Norway is in transition (Rindal, 2020) and as LK20 shows, students need to be capable of tackling the rapid changes and demands of a fluctuating knowledge society. This means equipping them with the strategies not only to communicate, but to engage with information critically and analytically, and to be able to learn even beyond the borders of the classroom. Moreover, teachers in Norwegian schools have considerable freedom and responsibility both methodologically and when choosing language content such as language models. Inquiry-based approaches to education seem well suited to meeting these methodological requirements, and data-driven learning as a mode of inquiry provides a unique approach to language inquiry that can introduce scientific ways of working with language that include multimedia texts and a focus on both form(s) and language awareness and variety. Data-driven learning as a mode of inquiry is the topic of Chapter 3.

3. Data-Driven Learning as a Mode of Inquiry: A Theoretical Framework and A Way Forward

3.1 Chapter Introduction

This chapter presents the theoretical framework of the dissertation. The aim of this chapter is two-fold. The first part of the chapter reviews the theoretical perspectives that have already been associated with DDL in the literature. These perspectives include discovery learning, constructivism, and socio-cultural theory. Several principles from these theories influenced the dissertation's research design and cross-references will be inserted along the way that point the reader to the appropriate sections of the methodology chapter (Chapter 4). In addition, it is argued in this part of the chapter that DDL has relied too much on its constructivist association where pedagogic implementations are concerned, which has created practical challenges and theoretical pitfalls such as an overreliance on students' researching capabilities, overly individualistic approaches to learning, a lack of classroom differentiation, and a vanishing teacher role. In order to meet these challenges and nuance our didactic view of DDL, new insights from *inquiry-based education* are proposed in the second part of this chapter as a theoretical contribution to both DDL and the English subject in Norway. The need for a pragmatic view of the aforementioned theories in their application to DDL, in which theories answer the issues or problems they are suited to answering, is thus argued for. Inquiry-based education builds on a socio-constructivist foundation (cf. Section 3.3), and shares many basic principles with DDL, but with a clearer undercurrent of socioculturalism from its conception (see Section 3.3.2). Moreover, inquisitive, scientific, and explorative approaches to learning are highlighted in the new core curriculum in Norway as conducive to in-depth learning (cf. Section 2.3.4) and this is further reflected in the competence aims of the English subject curriculum (cf. Section 2.3.5). It is argued in this chapter (1) that data-driven learning is a potential mode of inquiry – i.e., a way of doing inquiry in the English language classroom – that can help satisfy the aforementioned curricular ambitions, (2) that inquiry-based concepts and processes can offer valuable pedagogic insights through well-developed frameworks that focus on the facilitation and training of teachers and students to become practitioners of inquiry by increments and through a diversified view of role taking, and (3) that although DDL and inquiry already have intersecting ideas and ideals, some DDL approaches are not sufficiently founded in pedagogic practice and need to be re-examined or expanded upon.

Section 3.2 elaborates on the theoretical underpinnings of the dissertation's research and DDL in general as it is paramount to the discussion of DDL as classroom practice and DDL as a mode of inquiry. Section 3.3 presents inquiry as a potential way forward for DDL. This section includes how inquiry connects to the current research and the curriculum, its basic concepts and how they align with those of DDL, and its theoretical and practical contributions to DDL. Section 3.4 summarizes the chapter.

3.2 The Theoretical Foundations of Data-Driven Learning

Although corpora do not "initiate or direct the path of learning" (Leech, 1997, p. 5), DDL builds on assumptions about what languages are and how they are learned through the usage-based model (Section 3.2.1) and has been linked to several theories and hypotheses that influence how it is approached in the classroom and anchor it in contemporary epistemologies (Section 3.2.2). These theories conceptualize knowledge and learning and therefore influence the way different teachers and students engage in the educational process and what is valuable to them in terms of input and classroom practices, while they at the same time may restrict pedagogic practice by, for instance, focusing too much on the individual and not on the social process of education. The research design of this dissertation built on many of these theoretical perspectives concretely. For instance, the socio-cultural principles of learner agency, collaborative dialogue and metatalk were pivotal features in the planning and implementation of the classroom case study (see Chapter 4). Other perspectives underlie DDL approaches tacitly as assumptions about what elements are conducive to language acquisition and cultural learning or how teachers' epistemic beliefs impact their decision-making (cf. Article 1). The following sections elaborate on the theoretical underpinnings of DDL, how adopting them might steer educational practice in certain directions, and the contributions and shortcomings associated with their epistemic foundations. Connections to theories of language and language learning through the usage-based model specifically are presented first (Section 3.2.1), followed by DDL's links to broader learning theories such as constructivism and socio-cultural theory (Section 3.2.2).

3.2.1 DDL, Language in Use and the Impact of Instruction

A basic premise of DDL is that language learning is usage-based. According to the usage-based model, the majority of language learning happens as an implicit process through usage, but instruction can in some cases be valuable (Ellis & Wulff, 2015). Thus, establishing a pedagogy within a usage-based framework will have consequences for language learning and teaching, particularly in regard to the impact of direct instruction and learner engagement. In the following, the basic concepts of the usage-based model and their connection to corpora are covered before their pedagogic and practical implications are discussed.

The usage-based model posits that the language system is comprised of conventional form-meaning pairings (Evans & Green, 2006, p. 136) and not governed by underlying, innate rules. According to Ellis and Wulff (2015), these pairings – or *constructions* – accumulate in the speaker's mind as a "warehouse" of constructions that comprises one's language knowledge. The utterance, understood as a culturally- and contextually-embedded instance of language use, is central in the usage-based model; knowledge of language comes from language use; the interactive and contextual nature of human language is emphasized; and "the relative frequency in linguistic units affects the nature and organization of the language system" (pp. 146-147). These principles entail that language learning is a case of "[…] the learning of many tens of thousands of constructions (words, morphemes, lexico-grammatical patterns, etc.) and of the probabilistic relations between them and their functions, their speakers, their contexts, and their genres" (Ellis, 2019, p. 49) and is exemplar-based, frequency-based, and associative (Ellis & Wulff, 2015). What this means is that exemplars of constructions that are more frequent in the language we are exposed to are more easily processed, and that our perceptual systems become more attuned to items that are more likely to appear in the input (i.e., associative learning). Adopting a usage-based conception of language acquisition impacts education because it puts the usage event at the forefront and explains language acquisition as something beyond habituation and transfer (e.g., Bloomfield, 1935) or nascent syntax where language production – or use – is devalued (e.g., Chomsky, 1953). The focus on frequency and language in use resonates with the affordances of corpora, and it is perhaps unsurprising that corpus data are used as evidence for the usage-based nature of language learning, since corpora are, in a sense, a warehouse of authentic utterances. Moreover, corpora through their metadata can provide information about genre, demographics, and context, which coincides with the descriptions of language and language acquisition in the usage-based model, particularly exemplar- and

frequency-based learning and the contextual nature of language acquisition. Thus, although corpora are not inherently pedagogic, their connections to a usage-based model show their potential impact on practice. This impact is tied to the value of instruction and nature of language learning, which are considered next.

The pedagogic value of corpora, and language teaching/instruction in general for that matter, beyond just providing more exposure, should be considered in light of whether or not explicit instruction or teaching is useful and to what extent. This issue is a question of *knowledge interface*, or "[…] whether explicit and consciously taught (or learned) knowledge can ever be internalized by learners to become part of the implicit *automatized* sub-conscious knowledge system" (O'Keeffe, 2020, p. 8). According to O'Keeffe (2020), teacher-mediated approaches to DDL (see Section 3.2.2) align more with the *strong interface position*, which maintains that explicitly taught knowledge can become implicit knowledge, while DDL as discovery learning (see Section 3.2.2) aligns better with the usage-based model and the *weak interface position*, which posits that incidental noticing (discussed below) may lead to incremental acquisition of target forms (p. 9). These are positions where form(s) can be presented and practiced (strong position) or given additional attention (weak position) (ibid.). For instance, teachers that believe in the strong position might find direct instruction on specific forms conducive to their learners' uptake of these forms. DDL has predominantly been connected to the weak position through concepts such as Bernardini's (2004) *serendipitous learning*, where learners acquire an aspect of language by happenstance, while their focus is elsewhere. It has more prominently been associated with Schmidt's (2001) *noticing hypothesis* (e.g., Flowerdew, 2015; O'Sullivan, 2007), according to which language forms can be acquired if conscious attention is given to them. Examples of how corpora can facilitate additional attention and noticing are *enhanced input* and *frequency data*, which are arguably two of corpora's primary pedagogic utilities. For instance, concordance lines are usually sorted in a column for vertical reading and words are sometimes labeled or color-coded according to their grammatical category. In addition, both concordance lines and frequency lists can make explicit the most frequent words or keywords in the corpus, and thus draw learners' attention toward commonalities and idiosyncrasies in any represented genre or type of speaker. If one considers the utility of corpora as rich databases of authentic data that provide the learner with exposure and increased salience in light of the type of learning conceptualized in the noticing hypothesis, it makes sense to facilitate learner-to-corpus interactions as much as possible. While such interactions can be a gateway to varied and enhanced language input, they can also lead to individualized and cognitively demanding

working methods, especially if they are accompanied by an over-zealous view of discovery learning and student-centeredness.

The adoption of a usage-based framework and the weak position has further implications for pedagogy, and particularly the teacher's role. For instance, Ellis and Wulff (2015) point out that "[…] the bulk of language acquisition is implicit learning from usage" (p. 89), but for second (or foreign) language acquisition, "additional attention" can be necessary when the "linguistic form lacks perceptual salience [or] the L2 semantic/pragmatic concepts […] are unfamiliar" (ibid.). For instance, interactional partners or an instructor can direct learners' attention to specific forms and by doing so "[…] recruits the learner's explicit conscious processing" (p. 83). In other words, the teacher can facilitate exposure and enhanced input, and draw the students' attention to certain linguistic aspects, but the value of form-focused instruction is limited. There is therefore a vagueness in what the teacher's role and conduct in the language classroom actually consist of. This vagueness is exacerbated when one considers that noticing is not the end of the learning process, but an initial stage that provides an interface between input or exposure and the development of constructions (Schmidt, 2001, p. 31). That is, one only notices surface structures in instances of language, and not abstract rules or principles of language (p. 5). Consequently, instances of deeper learning involve higher-order cognitive skills such as induction or self-regulation, which will be discussed in the following section in connection to constructivism and socio-cultural theory (see Section 3.2.2), while the noticing hypothesis leaves us with an incomplete picture of language-learning processes.

The fact that DDL and corpus linguistics are so closely connected to the usage-based model and concepts associated with the weak interface position, such as noticing, may explain in part the link between DDL and cognitive constructivism and consequently how learner-to-corpus interactions are highlighted while teacher-learner relationships are overlooked (cf. Section 3.2.2). From a purely language-learning perspective, if acquisition is implicit and the concordancer can provide enhanced input, the teacher is left a space that is vague and undefined. Moreover, the English subject in Norway is about more than learning form-meaning constructions (see Chapter 2). The subject also focuses on culture, different competences (e.g., the ability to learn or communicate), and language-learning *strategies*, none of which enhanced input or additional attention can provide. A corpus-based approach must take the multifarious nature of the English subject – and what it means to be an English teacher and an English student – into account if its integration is to be successful and fruitful. In the following section, the links between DDL and constructivism and socioculturalism are examined in order to identify the

pedagogical principles that underlie data-driven learning and the implications of these principles.

3.2.2 Socio-Cultural Theory and Constructivism in Data-Driven Learning

The discussion of learning theories in the DDL literature explicitly references principles of constructivist theory and socio-cultural theory and conceptualizes learning with reference to these paradigms (e.g., Chambers, 2019; Flowerdew, 2015; O'Keeffe, 2020). The primacy given to a set of constructivist principles (see below), however, has arguably led to certain manifestations in practice while inhibiting or obscuring other opportunities. This choice of focus particularly impacts the nature of educational roles – i.e., how teachers and students should act in the classroom. The mixture of principles from both theories is not in itself problematic, as "stripped of their essentials, constructivism tells us to pay close attention to the mental activities of the learners, and socioculturalism tells us to pay close attention to cultural practices in the learner's milieu" (Bereiter, 1994, p. 21). In this pragmatic view, one needs "[…] to consider what various perspectives might have to offer relative to the problems or issues at hand" (Cobb cited in Bereiter, 1994, p. 21). In other words, the two theories are not mutually exclusive nor contradictory but can offer different perspectives on learning that need to be evaluated critically as they impact the classroom differently through teachers' beliefs, school policy and educational research. In the following, the principles of each of the theories are covered in terms of how they relate to the dissertation and DDL, and the pedagogic consequences associated with adopting these theories are discussed. Constructivism will be covered first, as it holds a particularly central position in DDL. Socio-cultural theory is examined second. Note that it is in the sometimes-subtle interface between these theories and how their principles are highlighted in different educational situations that the basis is formed for a sought-after theoretical pragmatism in the conceptualization of DDL, and this is subsequently where we find the foundation for rethinking DDL as a mode of inquiry (see Subchapter 3.3).

According to O'Keeffe (2020), "the associative link to constructivism is seen as a pedagogical hallmark for DDL" (p. 3). In constructivist learning, "students apply their general cognitive problem-solving mechanisms and existing background knowledge to foster an understanding of new data" (Flowerdew, 2015, p. 18). This means for DDL that the students engage with corpus data to construct mental representations of the language (cf. Section 3.2.1 for an account

of constructions in the usage-based model). Constructivist learning is, in practice, associated with inferencing, hypothesizing, student-centeredness, and discovery learning (O'Keeffe, 2020, p. 3) and inductive learning represents a core concept in both constructivism and DDL (Flowerdew, 2015, p. 29). Inference entails drawing a conclusion from the available evidence or data, while induction, similarly, is to discover principles or rules from data or evidence – i.e., the learning is *driven* by the examination of data. Construction of knowledge structures can thus be facilitated by corpus data through research-emulating activities (e.g., Johns, 1991), which include "predicting, observing, noticing, thinking, reasoning, analyzing, interpreting, reflecting, exploring, making inferences (inductively or deductively), focusing, guessing, comparing, differentiating, theorizing, hypothesizing, and verifying" (O'Sullivan, 2007, p. 277), and accessing data in different ways through a concordancer (Flowerdew, 2015). These principles align with some of the new ambitions of discipline-specific competencies and in-depth learning in the new core and English subject curricula in Norway but at the same time run the risk of being highly individualistic and not directly relatable to other aspects of the curriculum such as communication and participation (cf. Chapter 2).

The integration of the aforementioned constructivist principles in the classroom has impact on pedagogy that, on the one hand, has been held up as largely positive changes from traditional teaching in much of the literature, while on the other hand poses certain challenges and may have unfortunate effects (see Article 3). For instance, a constructivist epistemology enables the rejection of an objectivist epistemology, and consequently problematizes traditional teaching approaches. In objectivism, students are perceived as passive objects to which knowledge can be transferred, and from this perspective, there is little incentive to spend time on, and invoke chaos through, student-centered and student-active approaches if knowledge can just be passed on through orderly lectures. This is because objectivism posits the existence of an objective reality for learners to absorb, and "students are not encouraged to make their own interpretations of what they perceive; it is the role of the teacher or the instructor to interpret events for them" (Jonassen, 1991, p. 10). Meanwhile, constructivism is about how meaning-making relies on the individual's interpretations and existing knowledge structures (ibid.). In this view, it is absolutely necessary that the students make use of their cognitive mechanisms and prior knowledge to interpret and analyze information to construct new knowledge, and this is best achieved through student-active learning. These changes in fundamental epistemic beliefs open up for student-centeredness and simultaneously rejection of old traditions where the teacher is

the absolute expert and learning is a form of transfer of (pre-packed) knowledge from teacher to student.

While DDL scholars have in large part championed the above-mentioned changes (e.g., Boulton, 2010; Gilquin & Granger, 2010; Johns, 1991; Leech, 1997), such transformations of pedagogy are not without consequences. Flowerdew (2015) recognizes that constructivist principles pose high cognitive demands, and that open exploration and discovery are not suitable for everyone. In addition, the teacher-student relationship in constructivism has been criticized. According to Biesta (2016), a critical voice against constructivism, constructivist pedagogy views learning as *immanent* in that it becomes about bringing something out that is already inside the learner, while the teacher role vanishes (pp. 46-47). The constructivist paradigm, says Biesta (2016):

> [has] shifted attention away from the importance of *relationships* in educational processes and practices and has thus made it far more difficult to explore what the particular responsibilities of and tasks of educational professionals such as teachers and adult educators are. (Biesta, 2016, p. 63; emphasis in original)

Instead, he argues that the teacher is an *Other* that comes from the outside and brings something radically new to the student (pp. 48-49). Meanwhile, he rejects the "student-consumer whose needs need to be met in the most effective way" and advocates for the student to be open to unwelcome truths and not just 'learning from' but being 'taught by' the teacher (Biesta, 2016, p. 58). In the former, the teacher is a resource and the student controls what is learned, while the latter involves something entering the student's reality from the outside – an interruption – and the student must be willing to give such interruptions authority (Biesta, 2016, p. 57). Biesta's rhetoric is too dismissive of the constructivist paradigm, and one could argue that learning does not need to involve radically new information but also minor changes, confirmations, or nuances. Moreover, constructivism does not argue that knowledge or learning strategies are innate, but rather that they are (partially) built through general cognitive mechanisms. Nevertheless, the issues of the vanishing teacher role and the lack of focus on teacher-student relationships that Biesta points to are important to be aware of in DDL so as to avoid underestimating or undermining the teachers' role or creating a learning approach that focuses largely on the individual's engagement with the tool and task in a subject that is founded on communicative principles (cf. Section 2.3.2). This issue of individualization of the learning process is the impetus for the theoretical pragmatism posited in this dissertation that takes the social dimension of education into account.

Although Vygotskyan socio-cultural theory has also been connected to DDL, it has been a less frequent occurrence. In fact, it would appear that socio-cultural principles have been added to DDL as a response to problems of different learner preferences and high cognitive demands that come with a purely learner-to-corpus focus, as opposed to DDL being rooted in socio-cultural principles. The foundation of socio-cultural learning is the idea that people utilize existing, and create new, cultural artefacts (e.g., language, literacy, categories, etc.) in order to regulate, monitor and control their behavior and "[…] developmental processes take place through participation in cultural, linguistic, and historically formed settings […]" (Lantolf, Thorne, & Poehner, 2015, p. 207). A handful of central socio-cultural concepts have been held up as helpful in framing DDL beyond an idealized and individualized inductive approach (see Chapter 4 on how they were implemented in the dissertation's research). Among these are *collaborative dialogue* (Flowerdew, 2015) and *learner agency*, *mediation,* and *self-regulation* (O'Keeffe, 2020). *Learner agency* involves the students possessing greater control over their learning process as opposed to being passive recipients of teacher-transmitted information (O'Keeffe, 2020). In practice, this point coincides with the student-centeredness of constructivism but focuses less on the individual. However, the same pedagogical argument remains, namely that learning is not a case of knowledge transfer in objectivist terms, but a facilitation or mediation of student-active forms of learning.

Mediation is a particularly central concept in socio-cultural theory that has been touched on in DDL. Mediation refers to how people use symbolic artefacts as buffers to mediate the relationship between themselves and the socio-material worlds (Lantolf et al., 2015) and foremost among these artefacts is language. In DDL, it is proposed that language learning can be mediated through *metatalk* and *collaborative dialogue* (Flowerdew, 2015). The former denotes the use of language to discuss language, and particularly metalanguage in this case (e.g., linguistic categories such as nouns or adjectives). The latter denotes how student peers, or the teacher, can engage with each other dialogically to, for instance, formulate queries or describe language phenomena. Through these forms of mediation, one can start to imagine ways in which peer-to-peer and teacher-to-student interactions become central parts of corpus-based approaches to language learning. More recently, *self-regulation*, a central form of mediation, was brought up in conjunction with DDL (see O'Keeffe, 2020). Regulation describes a trajectory of control people can achieve over their activities and is divided into *object-*, *other-*, and *self-regulation* (Lantolf et al., 2015, p. 209). Object-regulation is "when artefacts in the environment affords cognition/activity" (ibid.). These artefacts can be reference works or, as in

the case study (see Articles 2 & 3; Section 4.4.3), task-guides and the concordancer. In other-regulation, mediation happens in conjunction with other people through feedback, guidance, etc. Lastly, *self-regulation* "refers to individuals who have internalized external forms of mediation for the execution or completion of a task" (Lantolf et al., 2015, p. 209). Self-regulated students have gained greater control of their own learning processes and are more self-sufficient. Socio-cultural theory therefore provides one way in which one can conceptualize the road toward more independent learners and implement DDL in less individualistic ways. In addition, the theory shows how learning relies on the people and objects with which we engage.

Finally, I return to the aforementioned theoretical pragmatism involving these theories to consider briefly how each can be related to DDL based on the issues or problems at hand. O'Keeffe (2020) suggests that DDL can be placed along a cline ranging from (1) constructivist, discovery focused with no curation of data, no target form, and no pre-instruction on form, but with several entry points into the data for learners, to (2) socio-culturally focused, with self-regulated, teacher- and peer-mediated learning, curated data, and target form(s) with pre-instruction on said form(s), as well as peer-to-peer learning (p. 6). This proposed cline is helpful in that it opens up for conceptualizing and building DDL activities and defining the educational roles within it in ways that are more versatile, diverse, and rooted in theoretical pragmatism, which is in line with the didactic framing of DDL as a mode of inquiry (see Section 3.3). Several of the principles were directly implemented in the research design of this dissertation – the case study in particular (see Chapter 4). Similarly helpful continua of DDL practice have been proposed. For instance, DDL can range from deductive to inductive (Liu & Lei, 2017; O'Sullivan, 2007) and from teacher-led to student-centered (McEnery & Xiao, 2011; Mukherjee, 2006). Although these continua signal diverse opportunities for DDL implementation, the incremental nature of the process or the nodes in-between the extremes on the above-mentioned continua are not investigated to a notable degree (see Kennedy & Miceli, 2016 for one practical exception). The idealized form of DDL appears to be when a learner-independent, inductive model is adopted, but it is not given that this idealized form is necessarily the most worthwhile or that it should shape our perception of DDL. Looking at DDL in a variety of ways may be the key to its successful integration into pre-tertiary education. This is not to say that that anything goes when taking a pragmatic stance, as theoretical pragmatism in this context should be taken to mean that practical approaches are well-founded in theoretical positions *as they contribute to the particular situation*. The following subchapter argues for the

conceptualization of DDL as a mode of inquiry, which encompasses both language learning and other central competencies of the new curriculum in a more holistic framework.

3.2.3 Section Summary

This subchapter discussed theoretical perspectives and language-learning hypotheses with which DDL is associated and on which the current research was built (see Section 4.4.1). It explored the concepts of constructivism and socio-cultural theory as well as the usage-based model of language learning to better understand the theoretical framework of DDL, its assumptions about language, and the practical-pedagogical consequences that arise from these ideas and impact the applications of corpus-based approaches in positive and negative ways. These concepts can offer ideas on how to structure tasks and activities in the classroom, for instance by striving for student agency and activity, but adopting only a few of them, for instance inductive learning, may lead to unfortunate side effects such as the vanishing teacher role or highly individualized learning processes. The next subchapter deals with the concepts of inquiry, its contributions, and how it can offer a potential way forward for DDL in pre-tertiary education.

3.3 Inquiry-based Education: A Way Forward

3.3.1 What is Inquiry? Terms and Origins

The following subchapter explores inquiry-based education and how DDL can be a mode of inquiry that relies on a theoretical and practical pragmatism to meet the complexities of the English classroom. Three elements discussed in the inquiry literature are covered in particular, namely the social learning aspect of inquiry (Section 3.3.2), the incremental aspect of inquiry (Section 3.3.3), and the educational roles of inquiry (Section 3.3.4). The terms *inquiry-based learning* (Blessinger & Carfora, 2014), *inquiry-based teaching* (Keiser, Burrows, & Randall, 2014), *inquiry-based teaching and learning* (Saunders-Stewart, Gyles, Shore, & Bracewell, 2015), and *inquiry instruction* (Aulls, Tabatabai, & Shore, 2016) have all been used to refer to the same basic principles of inquiry in education, albeit with a focus on different (or several) actors. In the following, the term *inquiry-based education* [IBE] or simply *inquiry* is used to encompass all of these terms in a broad sense that includes its multiple actors, contexts, and

principles; however, when citing specific works, the other terms are used in accordance with how they were denoted in their particular work in order to preserve the authors' intent and specificity. It is argued in the following that the challenges and shortcomings in DDL's practical implementations and theoretical foundations can be met by establishing an inquiry framework in which DDL can be placed that is also congruent with the new curricular developments in Norway (cf. Section 2.3.3). This subchapter asserts that the research on corpus-based pedagogies can greatly benefit and learn from an IBE framework. The features of inquiry and DDL are juxtaposed to make visible (1) their considerable overlap and aspects of DDL that should be continued in an inquiry framework, and (2) the way the *social dimension* permeates inquiry and flows into its conceptualizations of education (see Section 3.3.2).

Historically, IBE has its roots in the K-12 science education in the United States of America. According to Aulls et al. (2016), "inquiry and inquiry instruction have been part of substantial teacher-education curriculum revision, especially but not exclusively in the sciences" (p. 1). As Chapter 2 shows, similar inquiry-based and discipline-specific principles are featured in the current Norwegian curricular reform. The origins of IBE can be traced to the US' National Research Council [NCR], but IBE has been explored in more general terms and as a viable option in the arts, humanities, and social sciences (see Blessinger & Carfora, 2014), hereunder language education (e.g. Caputo, 2014; Franc & Morton, 2014). This origin means that some of the research and theoretical perspectives examined in the present chapter have been conducted and developed with a basis in principles from the sciences and later adaptation to language education. It is therefore pertinent to be cognizant of the discipline-specific differences in practices, traditions, methods, and types of data in the science subjects when compared to the language subjects (see Section 3.3.2). In other words, the intent of this chapter is not to suggest a one-to-one transfer or mapping of the principles of inquiry-based education, as it is presented in science education, onto those of DDL; instead, the aim is to learn from IBE and see what parts of it can function as a blueprint for DDL practice. Thus, looking at IBE in relation to DDL enables us to share experiences and principles between two or more disciplines so as to promote cross-disciplinary perspectives and avoid having the same discussion twice. It is noteworthy that there appears to be no mention of corpora in the IBE literature nor any mention of IBE in the DDL literature[2] despite their perceived compatibility (cf. Section 3.3.2).

[2] This statement is based on the literature reviews in Chapters 1 and 3. These were extensive, but might not be exhaustive, which means that DDL/inquiry crossovers might exist unbeknownst to the author.

Section 3.3.2 presents the central tenets of inquiry and how they align with DDL and language study, and the new curricular developments in Norway. In doing so, the social dimension of IBE is emphasized. In Section 3.3.3, different types of inquiry are discussed and an incremental approach to inquiry is investigated. Section 3.3.4 examines inquiry through new and diverse roles and activities. The chapter is concluded in Section 3.3.5.

3.3.2 Inquiry and DDL: Characteristics, Alignments, and the Social Dimension

The following section examines the characteristics of IBE, how they align with those of DDL, and highlights the social aspects that permeate IBE. One aim is to show how there is a social dimension of IBE that is foundational to the approach and that trickles down to the students' activities and teachers' conduct in the classroom (cf. Sections 3.3.3 & 3.3.4). The pervasiveness of the social dimension in the very conceptualization of IBE (see below) is a key feature that can arguably benefit DDL, where socioculturalism appears to have been brought up mainly to answer criticism of cognitive demands and individualizing approaches (cf. Section 3.2.2). In fact, inquiry is explicitly and consistently defined as a *socio*-constructivist approach (Aulls et al., 2016; Caputo, 2014; Keiser et al., 2014; Saunders-Stewart et al., 2015; Walker & Shore, 2015). In addition, another aim of this section is to show how DDL and IBE overlap, as these are elements of DDL that should be kept and built upon.

When comparing the characteristics and goals of DDL and IBE, several common traits emerge. At the heart of much IBE research and theorizing lies the definition of inquiry[3] given by the NCR:

> Inquiry is a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results. Inquiry requires identification of assumptions, use of critical and logical thinking, and consideration of alternative explanations. Students will engage in selected aspects of inquiry as they learn the scientific way of knowing the natural world, but they also should develop the capacity to conduct complete inquiries. (National Research Council, 1996, p. 23)

[3] Note that this definition lays the foundation for a particular direction of inquiry. Other types exist with different foundations and will be discussed in Section 3.3

In DDL, Callies (2019), drawing on the work of Mukherjee (2006) and Dalton-Puffer (2014), points to components teachers need to master in order to achieve *corpus literacy*. Among these components are an understanding of the affordances and concepts of corpora, and the ability to search corpora to analyze and interpret the data and extrapolate general trends/patterns from one's findings (p. 247). These skills are required for students as well (see Lee, Warschauer, & Lee, 2020, p. 346), if the goal is to "cut out the [teacher as a] middleman" (Johns, 1991) and give students direct access to corpus data. Corpus-based approaches thus require one to formulate queries or language hypotheses and make appropriate searches and analyses in order to answer them. It is central to both IBE and DDL that students should engage with research-like activities, and both approaches appear to share an ambition to make students autonomous practitioners of research-like learning, which is reflected in the NCR's (1996) statement about inquiry, "[students] should develop the capacity to conduct complete queries" (p. 23) and frequently mentioned in the DDL literature as a positive outcome of DDL (e.g., Crosthwaite, 2020a; Leńko-Szymańska & Boulton, 2015).

The call for more inquisitive and explorative approaches matches the ambitions outlined in the new core curriculum in Norway as well (cf. Section 2.3.3). In the Norwegian context, Andersen, Fiskum, and Rosenlund (2018) connect the focus on exploration, creativity, and engagement in the new curriculum (cf. Chapter 2) to promoting inquiry, a sense of wonder, and active participation in the classroom. They argue that learning activities that create *a sense of wonder* are supposed to contribute to curiosity and to make students pose more questions than they answer. *Inquiry*, on the other hand, is about students examining and answering a problem, hypothesis, or research question. In inquiry, the teacher sets the framework and goal, while the students choose the strategies and methods. *Active participation* coincides with the concept of student-active, student-centered learning. These descriptions share commonly recognized features of contemporary education such as learner empowerment and autonomy through learner-centered education, a focus on lifelong learning, and inductive learning principles, which resonate with DDL. The adherence to learner interests and curiosity represents both a student-centered ideal and a motivational factor in IBE that have been argued for in DDL as well (e.g., Bernardini, 2004; Hasselgård, 2014). The shared practical and epistemic principles founded in research-emulating and constructivist ideas outlined above are what make the comparison between IBE and DDL relatively frictionless; however, while DDL has largely relied on constructivist principles and only partially touched on socio-cultural ones (see Section 3.2.2), IBE is consistently and explicitly linked to *socio*-constructivism. These socio-cultural

principles tie IBE closer to the nature of the English subject and its roots in communicative competence and participatory learning (cf. Chapter 2).

The NCR principles described above have germinated into a variety of different definitions in more recent papers where constructivist and research-emulating techniques, such as induction and inference, are emphasized; however, one can clearly see references to the social dimension throughout (highlighted in italics below). Caputo (2014) defines IBE as:

> […] an exploratory, learner-centered teaching method that encourages the employment of *active student involvement*[4], inductive learning techniques, problem-solving activities, and the use of questions as a research foundation that can then be transformed into a basis for student inference and critical evaluation. (p. 370; my italics)

In addition, Caputo (2014, pp. 370-371) points to aspects such as lifelong learning, meta-cognitive awareness, learning skill development, critical thinking, process-oriented activities, *active students*, and exploration and discovery, as well as requirements such as risk-taking, *collaborative learning*, and self-reflection as central components of IBE (pp. 370-371; my italics). Similarly, Blessinger and Carfora (2014) describe IBE as "a *learner-centered, learner-directed,* and inquiry-oriented approach to learning that puts more control for learning with the learner" (p. 5; my italics) with the goal of "[…] developing more self-sufficient lifelong learners" (p. 8), and which "in addition to acquiring established knowledge, encourages students to construct their own new knowledge and *share that new knowledge with their peers*" (p. 13; my italics). In this definition, the inquiry-learning process does not stop with the construction of new knowledge through inductive learning but includes peer-to-peer interactions. According to Blessinger and Carfora (2014), inquiry-based learning centers around three main constituents:

1. Exploration and investigation (e.g., problem-based learning, *collaborative learning*, self-directed learning, meaningful learning),
2. Authentic inquiries using contextualized and situated learning (e.g., field learning, service learning, case-based learning), and
3. [a] research-based approach (e.g., research-based learning, *project-based learning*[5], *scaffolded learning*). (p. 14; my italics)

[4] Denominators such as *student active involvement, student-centered,* and *active students* can be interpreted as both the student-centeredness suggested by constructivism and the student agency suggested in socio-cultural theory (cf. Section 3.2.2).

[5] Project-based learning can technically be individual; however, it is often group-based and conducive to peer-to-peer collaboration and co-learning.

The above constituents tie IBE to several learning approaches, many of which align with DDL, like research-based learning and authentic inquiries, but that also bring further attention to the social dimension and the teacher role through scaffolded learning. Furthermore, Aulls et al. (2016) describe *inquiry instruction* specifically, which appears to be directed at how teachers can facilitate students' active engagement, acquisition, and *self-regulation* of knowledge construction through an inquiry process (my italics). This process includes connecting disciplinary and interdisciplinary knowledge and is influenced by students' individual interests and previous knowledge. The notion of self-regulation, which has recently been mentioned in connection to DDL (cf. O'Keeffe, 2020; Section 3.2.2), further solidifies IBE's commitment to the social dimension of teaching and learning and puts the responsibility at least in part on the teacher to facilitate such processes (see Section 3.3.4 for a discussion of roles in inquiry).

Lastly, despite the aforementioned similarities, there are certain characteristics of IBE as an approach based in the natural sciences that diverge from a corpus-based approach to language education that one must be aware of. Two such characteristics I have named (1) *datatype* and (2) *the nature of the object of research.* In relation to the former, the NCR definition given above uses the wording "[…] reviewing what is already known in light of experimental evidence […]" (NCR, 1996, p. 23). What constitutes 'experimental evidence' is a debate in and of itself that is beyond the scope of the current discussion but suffice to say that the notion of 'experimental evidence' does not necessarily fit neatly with the datatype in a corpus. Corpus data may be the product of experimentation where for example learner texts have been collected after a specific intervention; nevertheless, in most cases, corpus data are a collection of pre-existing texts and not a product of experimentation. Common ground in the case of (1) can be found, however, if the perspective is lifted to *empirical data*; both approaches are rooted in the examination of empirical data – they are data driven – albeit not necessarily in the way these empirical data are derived. The latter issue – the nature of the object of research – concerns the NCR (1996) goal: "Students will engage in selected aspects of inquiry as they learn the *scientific way of knowing the natural world*" (p. 23; emphasis added). The issues of what the 'natural world' is and how it can be known are interesting ontological and epistemological questions that will not be answered here, but this quote clearly highlights IBE's position in a scientific paradigm. As Andersen et al. (2018) point out, certain school subjects – such as natural science – build on accepted scientific explanations and (neo-)positivism, while other subjects, like the languages, build on hermeneutic and interpretive principles where different interpretations, argumentation, and discussion are central elements (p. 18). Therefore, it is

important to emphasize the *social* and *mutable* nature of language as a phenomenon compared to the objects of research in a biology or physics class. Here, too, one can find common ground, however, in that corpus-based approaches to language education seek to provide students with an awareness of how languages are actually used through authentic examples (see Leńko-Szymańska & Boulton, 2015). At the center of both these approaches is an ambition to help students see a representation of 'reality' and to raise their awareness by investigating it. However, awareness of language as an object of research is not about understanding some underlying law or rule, but rather to understand its variations, socio-cultural situatedness, and adaptability.

In summary, IBE and DDL are approaches with considerable overlap both in practice and theory, but with some notable differences in their intended objects of study and in their focus on the social dimension of education. These commonalities and variations open the door for reciprocal influence albeit in different ways. In the following, different types of inquiry are explored in order to build a framework where DDL becomes a mode of inquiry with a view to incrementality and educational roles in particular. The social dimension of IBE explored in this section is a central component in achieving this framework.

3.3.3 Inquiry by Increments

Inquiry is defined through different terms and practiced through varying approaches that impact the degree of teacher and student involvement and engagement, as well as the openness and scaffolding of the learning activities and tasks. According to Saunders-Stewart et al. (2015), *open inquiry*, which entails a minimal teacher presence and near-absolute student autonomy, has been considered the gold standard; however, more teacher-directed approaches to inquiry have come to the fore. Minner, Levy, and Century (2010) performed a meta-study (138 analyzed studies) on inquiry instruction in K-12 science education and found a positive trend in its outcomes, but little evidence that open inquiry was a better alternative than more teacher-directed types. Moreover, Kirschner, Sweller, and Clark (2006) criticized what they termed *minimal instruction*, which included discovery learning, inquiry learning, problem-based learning, constructivist learning, and experiential learning. They concluded that minimal instruction based on the assumptions that (1) learning should happen through 'authentic' problems and (2) pedagogic content is parallel to the discipline's research methods, processes, and epistemologies does not work. Instead, they argued that "strong guidance" is more effective

based on both empirical evidence and our knowledge of cognition and that one should teach central concepts and the processes of the subject. In direct response to Kirschner et al. (2006), Hmelo-Silver, Duncan, and Chinn (2007) pointed out that the categorization of their review was imprecise, as at least problem-based learning and inquiry are not minimal instruction approaches, "but rather provide extensive scaffolding and guidance to facilitate student learning" (p. 99), and that inquiry learning and problem-based learning are well supported by empirical evidence (ibid.). These two approaches have in fact been scaffolded through (a) modeling and coaching with gradually less support, (b) prompts for strategy use, (c) teacher-provided structure through guides, fill-in activities, and diagrams, (d) templates and domain-specific explanations, (e) hypothetico-deductive approaches (see below), and (f) expert information and guidance granted to the students (Hmelo-Silver et al., 2007). Parallels can be drawn between the above-mentioned minimal instruction criticism, this dissertation's criticism of minimal guidance in DDL, and Biesta's (2016) criticism of the vanishing teacher in constructivism (cf. Section 3.2.2). However, this dissertation does not necessarily argue for a return to teacher-directed learning, but to a multifaceted approach along the lines of inquiry outlined in this subchapter. Importantly, Hmelo-Silver et al. (2007) position inquiry somewhere in between open discovery on the one hand and strong guidance on the other hand. These are conducive to an incremental approach to inquiry and subsequently DDL.

One can thus imagine different types of inquiry based on teacher involvement, scaffolding, and student responsibility and autonomy. Banchi and Bell (2008) conceptualized four levels of inquiry:

1. Confirmation inquiry: Students confirm a principle through an activity when the results are known in advance.
2. Structured inquiry: Students investigate a teacher-presented question through a prescribed procedure.
3. Guided inquiry: Students investigate a teacher-presented question using student designed/selected procedures.
4. Open inquiry: Students investigate questions that are student formulated through student designed/selected procedures. (p. 27)

In the first step (1) the question, procedure, and solution are provided by the teacher; in the next step (2), the question and procedure are provided; in the third step (3), only the question is provided; meanwhile, the final step (4) provides nothing (Minner et al., 2010). This hierarchy

is helpful in that it supplies us with a way of conceptualizing types of inquiry and a concrete incremental approach toward increased student autonomy and/or potential differentiation.

These types of inquiry can represent nodes along the deductive-inductive continuum that has been proposed in DDL (e.g., Lee et al., 2020; Section 3.2.1) and are one way in which DDL can become a mode of inquiry. Deductive and inductive learning are opposite poles on this continuum, while hypothetico-deduction provides a middle ground. In hypothetico-deductive learning, the student starts with some preconception such as a theory, rule, or statement from the teacher, and it is up to the student to evaluate these preconceptions based on either their own chosen methods or one provided by the teacher (Andersen et al., 2018, p. 23). Meanwhile, deduction coincides more with the notion of learning as transfer while induction relates to learning as discovery (ibid.). Arguably, perspectives that are more hypothetico-deductive coincide with socio-cultural principles in two major ways. First, they focus on providing considerable teacher scaffolding. Second, they rely on the examination of cultural artefacts such as language definitions, dictionaries, theories, hypotheses, etc. Conversely, inductive approaches where the students discover patterns based on data and previous knowledge are closer to constructivist principles. The first type – confirmation inquiry – has the intent of transferring a pre-determined language aspect and can be placed close to the deductive pole on the continuum. In DDL, confirmation inquiry could be the teacher asking the students to check the dictionary definition of some linguistic unit against the corpus data, which would give the students experience with the tools without the added demands associated with proposing hypotheses or inductively inferring patterns. Inquiry types two and three are more hypothetico-deductive or inductive, depending on the specificity and amount of information provided in the teacher's question. In line with these positions, one could imagine the learner as a hypothetico-deductive researcher. Lastly, the fourth type represents discovery and may be linked to DDL in its original and most idealized form where the students are self-sufficient, and the teacher is a facilitator of the learning process. These four types of inquiry can represent both a gradual movement toward more student-active and cognitively demanding forms of inquiry, and a form of differentiation where different students can be provided different amounts of information, tasks, and methods based on their performance. According to Fiskum, Myhre, and Rosenlund (2018), the transition to inquiry-based approaches from other types of learning entails a *novelty space*, or an unfamiliarity, that will vary from student to student and is best bridged through guided induction and with student variation in mind. The necessity for differentiation relates to

how well students adapt to new learning situations and is pivotal to the acquisition and enactment of novel educational roles (see Section 3.3.4).

The difference between discovery and inquiry learning is particularly central to the examination of DDL in this dissertation, as it has been a primary concept of learning with which DDL has been associated. While discovery learning entails independent exploration and a trial-and-error process, and is described as unstructured, inquiry is "a more-guided process based upon expertise from teachers and peers", which is also more *social* in its focus on collaborative learning and dialogue (Saunders-Stewart et al., 2015, p. 290). Discovery is, in this sense, close to Banchi and Bell's (2008) open inquiry (above). However, discovery and open inquiry seem to differ in that discovery is freely exploratory in a wider sense and the students are not assumed to possess researcher-like competence (see Bernardini, 2004), while open inquiry requires that students have acquired the competence and knowledge to formulate questions and select appropriate procedures. The categorical distinction between discovery and inquiry does not appear to be firmly established in all inquiry literature, however, as Caputo (2014) writes, "[…] the environment [of an inquiry-based learning classroom] facilitates exploration and discovery […]" (p. 371). Elements of discovery and exploration may therefore have a place in inquiry, but not as the prime driving force of the approach. Caputo (2014) acknowledges that inquiry "[…] is a *spectrum* of approaches, all of which are justified by common constructivist principles" (p. 371; emphasis in original). I would strongly argue that inquiry-based education is united by *socio*-constructivist principles that are anchored firmly in Vygotskian socio-cultural theory as well. The epistemic cornerstones and practical variations outlined in this section are the basis for an incremental and adapted approach to DDL as a mode of inquiry and are foundational to the construction of a framework of roles in DDL.

3.3.4 New Roles through Inquiry

The establishment of a student-centered DDL that leans heavily on a constructivist foundation has had positive outcomes relating to student empowerment and active engagement, but a too-idealistic version of it can have consequences for the educational process, especially when younger learners are concerned. These consequences are an overreliance on students' researching capabilities and the vanishing teacher role (see Section 3.1) that come with the notion of *role shift* (see below), as well as the focus on the individual in the educational process (cf. Biesta, 2016; Section 3.2.2) that runs the risk of overlooking the multifaceted and complex

relationships that exist and evolve between teacher and student or student and student, and is at odds with the communicative focus of the Norwegian curriculum (see Chapter 2). Among the contributions that accompany DDL as a mode of inquiry is a nuanced and diversified concept of roles teachers and students would have to take on in the classroom. The term *educational roles* is used in this context to denote the actions, interactions, and functions associated with the actors in the classroom, namely teachers and students, as well as their requirements, investments, and responsibilities in the educational process. These conceptions of roles find support in the social dimension of inquiry (cf. Section 3.3.2) and the incrementality and differentiation that come with adopting different types and approaches to inquiry (cf. Section 3.3.3). According to Harmer (2015), "[teachers] are called upon to assume a number of different roles in the classroom, depending on what we hope our students will achieve and also on what they actually do" (p. 116). In addition, student-centrism is a growing trend in today's educational policy (cf. Chapter 2). This section explores how inquiry challenges traditionally established role divisions and responsibilities, and consequently the structure of the classroom, while still preserving the role of the teacher. In DDL, this challenge to the established structure of traditional classroom roles is described as a *shift*. For learners, this shift is one from passive recipient of knowledge to active participant; for the teacher, the shift entails a de-emphasis of the teacher as lecturer/giver of knowledge toward the teacher as a director, coordinator, facilitator, etc. This section presents the potential pitfalls of dichotomizing role evolution as a shift from one state to another and instead relies on the concept of role diversification (see below) for a more nuanced and productive image of educational roles, role acquisition and role transition. Approaching new activities in increments (cf. Section 3.3.4) is intrinsic to the following conceptualizations of roles.

At the heart of the argument for new ways of thinking and structuring educational roles in DDL is the contributions of Walker and Shore (2015) (see also Article 3). They claim that roles in inquiry have been conceptualized as *shifts*; role shifts imply the abandonment of current, traditional teacher and learner roles and the adoption of new ones (p. 1), which is equivalent to the role discourse in DDL (see Chapter 1 and Article 3). Instead, a framework is suggested where the gradual adaption of new roles and role features lead to *role diversification* where "[…] several roles, including roles not traditionally ascribed to the individual, could potentially be adopted at one time" (Walker and Shore, 2015, p. 1). This concept of roles has several implications for practice. Firstly, teachers are not marginalized, as they would need to switch between several roles – even traditional ones. Secondly, greater care is taken to ensure gradual

role adoption through an incremental process that keeps student variations and differentiation in mind. Thirdly, the roles become more fluid based on what the situation requires, and roles are less defined as new features and traits are introduced. Consequently, the permanence of role shift instead becomes a *flux* of role taking. Fourthly, student-student and student-teacher dynamics would also fluctuate based on the needs of the student(s) at the time, the problem that is being worked on, and the relationships between these actors. Role diversification therefore requires pragmatism and negotiation.

Walker and Shore (2015) suggest a four-stage process of role diversification in inquiry for both teachers and learners (pp. 7-10). This process became the theoretical framework in one of the dissertation's articles (see Article 3) and is further elaborated on in the following. Their framework considered teachers and learners' attitudes, beliefs, norms, expectations, previous experiences, and social factors in its design, and proposed a trajectory of role acquisition that moved from *exploration* to *engagement*, then *stabilization*, and finally *diversification* as the ultimate stage. As the first stage, exploration means that both students and teachers start exploring the inquiry setting and "how role expectations in these environments might differ dramatically from role expectations in a traditional classroom setting" (p. 8). This stage is arguably where the students and teachers first encounter the *novelty space* (cf. Section 3.3.3) and must begin to explore what it means to work through inquiry. For instance, when transitioning into inquiry activities, students might need to "unlearn" certain fossilized learning patterns and this process will rely on the individual students' adaptability, interests, attention span, and relationship to their teacher (Fiskum, Thorshaug, & Husby, 2018). In the next stage, the students "begin to formally adopt and engage in an *inquiry-student role*" (Walker & Shore, 2015, p. 8). Students would now be actively involved in inquiry activities such as teamwork, taking initiative, communication, creativity, organizing information, generating questions, and interpreting data, but conflicts due to former roles are expected (ibid.). These conflicts may be due to the aforementioned novelty space or social factors such as a fear of exposure in a group setting, and the relative safety familiar learning patterns offer. When such conflicts are resolved and students and teachers are dedicated to the new approaches and activities, they have reached the fourth stage of *stabilization* and value the changes ushered in by inquiry (Walker & Shore, 2015, p. 9). The last stage is *diversification*, which "refers to a process of expanding the repertoire of roles that an individual adopts, in number and variety" (Walker & Shore, 2015, p. 4). In this stage, the teacher and the student can switch between different roles as is required by the situation and show a high degree of versatility in handling different learning environments

and tasks. For instance, students can take the role of teacher, team leader, hypothesizer, presenter, audience, or explorer (ibid.). Adopting such a diverse view of roles, the teacher can consider the situation at hand and see when and for whom it is necessary to pre-teach certain forms before the corpus is consulted, or take on a student role and learn the corpus tool alongside the student to model the learning process, etc.

The novelty of inquiry roles can be approached in different ways. For instance, both DDL and inquiry are in principle student-centered and would entail a change in the teacher's responsibility from a teller of information to a facilitator of learning. This change means that the teacher spends more time on quality task and lesson design and preparation; a teacher that models learning, for instance by solving a task in the classroom while thinking out loud; peer scaffolding; the nurturing of a more positive learning climate, and; a focus on self, peer, and student evaluation of student work (Weimer, pp. 72-84). The teacher must further facilitate a good learning environment, resources, and good tasks, as well as provide meaningful feedback and encourage participation, mutual understanding, and shared responsibility (Doyle, 2011, pp. 52-53). However, it also means that the teacher must be cognizant of how students respond to these new principles and consider if more structured or guided inquiry is necessary (cf. Section 3.3.3) or if pre-teaching certain forms would be fruitful (cf. Section 3.2.2). Moreover, the teachers should strive to make their expectations of the students predictable, adapt the difficulty of the tasks, lend extra support to some students, build good student-teacher and student-student relations, and familiarize students with the methods, techniques, and equipment of the novel approach (Fiskum, Thorshaug, et al., 2018). In addition, teachers must examine their beliefs about students and their epistemic beliefs, as they might be unaware of how these beliefs impact their practice and view of their students. For instance, do they value order in the classroom or accept the chaos of group work; do they see students as object to which information is transferred or subjects that construct knowledge in different ways; and do they perceive their students as frail or robust, which would determine to what degree they could handle novelty (Fiskum, Myhre, et al., 2018, pp. 35-37).

All of these aspects highlight the complexities of being a teacher and the necessity to see teacher and learner roles in terms of diverse role taking when implementing DDL. For instance, it would appear that DDL scholars have tended to see students as primarily robust and self-sufficient, while the teacher role has been marginalized. The need for an incremental and diversified approach to DDL is what is meant by DDL as a mode of inquiry. These perspectives are further discussed against the findings of the articles (Articles 1, 2 & 3; Chapter 5) in Chapter 6.

3.4 Chapter Summary

The first part of this chapter reviewed the theoretical foundations and associations of DDL, their impact on pedagogy, and identified both potential strengths and weaknesses in the adoption of a constructivist framework. The second part of the chapter built on these theoretical observations and obstacles, and proposed a new way forward through inquiry in which a renewed focus on theoretical pragmatism was proposed that, if manifested through the social dimension of IBE, opened the door for ways to conceptualize DDL as a mode of inquiry where the complexities and challenges of teaching and learning in the upper-secondary classroom can be met in various manners that promote differentiation, incrementality, and a diverse view of role taking. The concrete suggestions of how this theoretical deliberation and its impact on pedagogy can be handled in practice are further explored in the discussion in Chapter 6 where the findings and experiences from the dissertation's three articles (see Articles 1-3) are discussed in light of an inquiry framework. This discussion will lead to advice for teachers, learners, researchers, and teacher educators. As mentioned throughout the current chapter, several of these theoretical principles have impacted the research design of this dissertation (see Section 3.2.2 in particular). Thus, the overall methodology of the dissertation is presented and elaborated on in the next chapter with references to theory from the current chapter.

4. Methodology

4.1 Chapter Introduction

The current chapter describes the two research phases of the project, its participants, the data-collection methods, data analyses, and ethical consideration, and includes methodological reflections throughout. The goal is to show not only how the individual methods were used to answer the research questions of each article, but also how they are connected and how they contributed to answering the main research question of the dissertation: *How can corpus-based approaches be integrated into Norwegian secondary schools and how are they received by the users?*

The dissertation contains this introductory chapter and three articles based on two research phases. Table 2 outlines the two different phases of the research design, which took place about a year apart and drew on a range of different data-collection methods. The data accumulated during the first phase laid the groundwork for the first article of the dissertation, while the data accumulated during the second phase formed the basis for the second and third articles (cf. Section 1.4). The first phase explored beliefs and previous experiences of students and teachers related to corpora prior to researcher interference through an online questionnaire to students and research interviews with their teachers. This phase aimed to understand the state of corpus implementation in these classrooms and uncover potential barriers to and opportunities for future implementations. The second phase involved a two-week cooperation with one teacher and his two classes, where corpus-related applications were integrated into their English classes. Data from the second phase were collected through classroom observation, teacher/researcher correspondence, and group interviews with a selection of students.

Table 2
Methods employed at each stage of the study and the resulting articles

Research phase	Time frame	Data sources	Articles
1	Dec 2018	Interview (teachers) Questionnaire (students)	1
2	Nov/Dec 2019	Observation Group interviews (students) Teacher / Researcher correspondence	2 & 3

In Section 4.2, the participants and sampling strategy are described. Sections 4.3 and 4.4 cover the research design of phase 1 and 2 respectively, before Section 4.5 moves into the data analyses. In Section 4.6, the transferability and trustworthiness of the research periods are highlighted, while ethical considerations are discussed in Section 4.7. The chapter is concluded in Section 4.8.

4.2 Participants and Sampling

4.2.1 Participant Overview

The study targeted teachers and students at upper-secondary school in Norway. In the first phase, the sampling was restricted to two schools in Oslo and two schools in Inland Norway. Three of the teachers held master's degrees, while the fourth was in the process of obtaining one at the time of data collection (cf. Section 4.2.3). All of them were active in-service teachers in upper-secondary schools at the time of phase 1. Two of the teachers were women and two were men. As Table 3 shows, 154 students participated in the first phase of the project, while 69 students participated in the second phase; note that the students in phase 2 were different students from those participating in phase 1, but that the teacher, John, participated in both phases. The students of the first phase were divided among a total of nine classes, while the second phase followed one of the teachers and two of his newly-allocated classes the following school year. The students were first- and second-year students in the first phase and first-year students in the second phase.

Table 3
Overview of the teacher and student participants

		Phase 1				**Phase 2**		
School	Teacher	Class 1	Class 2	Class 3	**Total Phase 1**	Class 4	Class 5	**Total Phase 2**
W	Nora	16	21	-	**37**	-	-	-
X	John	29	30	-	**59**	36	33	**69**
Y	Marcus	20	5	6	**31**	-	-	-
Z	Sarah	8	19	-	**27**	-	-	-
Total					**154**			**69**

4.2.2 Sampling Strategy

The initial sampling of teachers was done through purposive sampling. In purposive sampling, "the researchers handpick the cases to be included in the sample on the basis of their judgement of their typicality or possession of the particular characteristic(s) being sought" (Cohen, Manion, & Morrison, 2018, p. 218). In this study, the teachers were selected based on two characteristic: (1) they were in-service English teachers in a Norwegian upper-secondary school and (2) they had had a course in pedagogically-oriented corpus linguistics as part of their teacher education. Meanwhile, the student participants were approached and selected through these teachers, who were encouraged to involve as many of their students as possible. According to Schreier (2018), within a purposive sampling strategy one can distinguish between homogenous and heterogeneous samples. The former is based on a likeness between the participants in the sample, while the latter seeks differences between the participants in an attempt to maximize variation (p. 88). The current study's sample is based on the criteria given above, or *criterion-based sampling*, which is a subcategory of homogenous sampling. Palinkas et al. (2015) state that "[homogenous approaches] are used to narrow the range of variation and focus on similarities" (p. 534). The likeness between the teachers' educational backgrounds opened up the opportunity to compare their experiences and discuss the influence a similar corpus background has had on their different teaching environments and styles. Additionally, by sampling participants only from upper-secondary school – the first criterion – it allowed for comparisons between similar curricular demands, which can be a source of teachers' practical choices.

The underlying goal of one's research represents another rationale for the sampling approach (Schreier, 2018, p. 89). As Thomas (2006) points out, "because statistical inference is not a viable option (or goal) [in non-probabilistic sampling strategies in qualitative research], the careful matching of strategy to purpose is critical in defining the value of the research design and eventual analysis" (p. 403). Since the goal of the study was to get insight into the practice of in-service teachers with corpus backgrounds, to see how their background had affected their students, and subsequently to cooperate with them in order to implement corpora in their classrooms, criterion-based sampling was a logical way of ensuring the right informants to achieve these specific goals. It can therefore be viewed as a selection based on a specific type of case; in this respect, contrast was not necessarily sought, but rather participants that held a certain background and special insights so as to serve as information-rich informants.

The connection of sampling strategy and research goal was not finalized when the teachers were first contacted but emerged and evolved throughout the process. This emergent approach relates to a third distinguishing aspect of purposive sampling put forward by Schreier (2018), namely whether the sampling strategy was made in advance or emerged during the study (p. 88). The latter, emergent approach is usually said to be more fitting to the emergent perspective of qualitative research (Schreier, 2018). The sampling strategy was adapted after examining an initial pool of teachers from both lower- and upper-secondary school before the abovementioned characteristics emerged and were ultimately chosen. This resulted in a shift in sampling focus from a more general focus on any secondary-school English teacher, toward a particular focus on upper-secondary-school teachers with corpus backgrounds. Thus, the sample was narrowed and homogenized to a greater extent than first envisioned. Consequently, this narrowing also entailed a modification of the research goal from a more general question about corpus literacy in secondary school, to examining the influence of corpus courses on teacher practice and subsequent student experiences.

It proved difficult to reach a large number of teachers. Since teachers are often approached to participate in research, they are sometimes cautious about how they spend their time. By engaging a personal network, an initial pool of teachers was found that further suggested other teachers that might participate. If more teachers could have been reached in the original sampling strategy, one might have found other teachers with corpus backgrounds situated in different school environments when the final, purposive sampling strategy was set in motion. Since qualitative research often involves data saturation as a sampling aim (Palinkas et al., 2015, p. 534), the lack of more prospective teachers to approach to ensure greater saturation is a limitation of the project's first phase. Perhaps if the research design had kept a focus only on the teacher or the student instead of engaging both, thus making participation less demanding, one might have found more individuals willing to participate. Furthermore, if the research had relied solely on the online questionnaire, it may have been easier to distribute to more institutions in order to achieve a higher response rate than in a study requesting interviews. Nevertheless, the research sought a dual perspective from those who would potentially be most affected by a corpus-based approach, so all participant perspectives were necessary.

As mentioned above, the students were asked to participate through their teachers and the teachers were encouraged to ask as many of their students of English to take part as possible. The low student numbers in Marcus's classes 2 and 3 (see Table 3) were due to those classes being small vocational-studies classes. All other classes, including Marcus's class 1, were some

form of general-studies classes. General studies is by far the most populated study direction in Norwegian upper-secondary schools (see Utdanningsdirektoratet, 2019). The programs the students belonged to were coincidental as far as sampling was concerned, as it depended on what programs the teachers happened to teach. This sort of sequential sampling design – i.e., where the students are sampled as a consequence of the teacher sample – is potentially problematic in that it restricts the second sample and the subsequent data analysis (Palinkas et al., 2015, p. 537). In the context of this project, the student sample becomes closely connected, biased, relatively small, and non-random. This fact has implications for the chosen mode of analysis, which will be discussed in more detail in Section 4.4. Having discussed the sampling strategy and some of its implications, the chapter proceeds to presenting the teachers and students that participated in the study, before describing the research design and methods of data collection.

4.2.3 Teacher Profiles

Nora

Nora spent a year in the US and began her higher education with Latin-American studies, but promptly switched to English. She explained that the language itself was as important to her as wanting to become a teacher. She has a bachelor's degree in English, has pedagogical qualifications, and at the time of this study she was doing her master's degree in English part-time alongside working as a teacher in upper-secondary school.

She emphasizes knowledgeable teachers and giving students a safe learning environment as important factors in English teaching and sees the increased focus on in-depth learning as a positive educational development. The idea of teaching as instructions/lectures that "plant something" in the head of the student and lets them leave the classroom with nothing more to learn, she describes as outdated; instead, she views education as a lifelong process.

The school she is employed at she describes as a bit Montessori-inspired. Concerning digitalization, she thinks the school was forward-leaning in the past when it came to purchasing digital tools such as computers, Macs, and smartboards, although the latter have scarcely been used. She thinks her colleagues have good digital competence, but it is not a big topic of conversation at her workplace.

John

John chose English as part of his general teacher training through what he describes as "a process of elimination". The choices were English, Norwegian or mathematics, and the latter two were less tempting. Language in itself has not been hugely important in his life outside of his everyday need to communicate.

He strongly emphasizes written and spoken communication as the most important aspects of the English subject in school, describing language as a tool to achieve communication. Central to communication, he says, is that the recipient of a message understands what the sender wishes to say. Conversely, having an accent, or pronunciation and intonation that are "a bit off", are less important aspects of language use, as long as they do not hinder communication.

The leadership at his school is very eager when it comes to digitalization, he explains, perhaps a bit too much so. All the students are equipped with computers, either from the school or from home. John still believes in the advantages of writing on paper, and he appears skeptical to certain new digital tools, some of which he describes as poorly designed.

Marcus

Marcus worked odd jobs after upper-secondary school and eventually studied economics for a year. Later, he took a bachelor's degree while working for an American firm writing articles for technology websites. To get formal qualifications confirming his English competence, he took a year of English in a teacher-training program, which eventually led to a course in pedagogy, and then a master's degree in English didactics. Digital technology has long been his hobby.

In the English subject, he particularly points to English as a global language and communication in context as central elements. He emphasizes the importance of shifting the focus from simply imitating British or American English toward different varieties, such as African English varieties. In addition, he expressed how he aims to give his vocational students English vocabulary they can actually use in their professions.

The digitalization at his school he describes as lagging behind. For instance, the students' computers are old, and the school invested in smartboards – an investment that would have been better employed going toward laptops and tablets, among other things.

Sarah

Sarah's road to becoming a teacher began with an interest in languages, especially English, but she had originally no plans of becoming a teacher. She worked for a while as a substitute teacher and the enthusiasm for the profession started growing from there. She has studied social anthropology in addition to her teacher education, so her interest in the subject lies more in language-as-culture than in linguistics.

She points to cultural and communicative foci as the essence of the English subject, and that the subject should help students communicate successfully. Her lessons are largely student-centered, and she expresses the importance of teachers understanding how students learn differently and how they learn through trial-and-error. She also thinks it is valuable to have the students learn together, something she described as "an awkward socio-cultural answer" in the interview.

The school she works at she describes as ahead when it comes to digitalization. She stated that everything at her institution happens on computers, and that there are very few course books or paper materials. In fact, her school has digital competence as a prioritized area of development. She experiences her colleagues as positive to digitalization in general and explains that the school even has two designated digital supervisors.

4.2.4 Student Profiles

As shown in Table 3, there were a varying number of students from each teacher in the first phase. Most of the students were in a general-studies program, but John's students in both phases had a particular specialization, while Marcus' second and third groups were vocational students. The choice was made not to disclose the particular specialization or vocation in order to prevent identifying characteristics and secure anonymity.

Gender was self-reported in the questionnaire in the first research phase. The gender balance across all groups were 71 female (46.1%), 78 male (50.65%), 2 students who answered *other*, and 3 students who did not wish to answer the question. In the questionnaire, students were asked to rate the degree to which they enjoyed English as a school subject, and how they assessed their own written and oral English skills based on the earlier feedback they had received from their teacher. The answer categories were Likert-type items of *strongly agree, partially agree, neither agree nor disagree, partially disagree,* and *strongly disagree* for the

former, and *very good*, *good*, *average*, *weak*, and *very weak* for the two latter categories. They were also asked whether they saw themselves choosing English going forward beyond the obligatory first year, with the answer categories *yes, I do not know, no,* and *I have already done it* – this final category aimed at the second-year students who were already taking additional English. These questions came at the very end of the questionnaire to avoid provoking students and color the other results. Out of 154 total respondents, 61 students (39.61%) strongly agreed, and 45 students (29.22%) partially agreed with the statement that they enjoyed English as a school subject. In other words, there was a positively skewed opinion of subject English among the participant students with a total of 106 responses (68.83%). Meanwhile, 29 students (18.83%) neither agreed nor disagreed with the statement, which left 19 students (12.33%) who viewed subject English negatively. Regarding whether or not they would keep studying English in upper secondary beyond the first year, 24 students (15.58%) answered yes, 49 students (31.82%) answered no, 52 students (33.77%) were undecided, and 29 students (18.83%) had already done so. The answers to these two questions paint an incomplete picture of the students' general motivation toward the English subject but indicate that they are not predominantly negative to English, which might have skewed the results of the questionnaire toward more negative responses. More information would have had to be gathered to judge their motivation, but these results give a general idea of their opinions going into the questionnaire.

The students' evaluations of their oral and written English skills based on previous teacher feedback were mixed but positively skewed. For oral skills, 33 students (21.43%) reported very good skills, 63 students (40.91%) reported good skills, 47 (30.52%) said average, and 6 (3.9%) and 5 (3.25%) students said weak or very weak, respectively. For written skills, 22 students (14.29%) considered their skills to be very good, 69 students (44.81%) answered good, while 46 students (29.87%) though they were average based on teacher feedback. Meanwhile, 12 students (7.79%) reported that they were weak in written English and 5 students (3.25%) responded very weak. The mostly positive answers on these two items lend support to the claim that Norwegians are considered competent English users (Brevik, 2019; see also Section 2.2; Crystal, 2012), but one should keep in mind the variation and the need to differentiate in order to reach every learner.

The students of the second phase were asked to fill in a short information form about themselves prior to the case study implementation. They reported information about gender, age, whether they had lived in an English-speaking country and, if so, for how long, and what language they knew. Out of the 69 students that participated, 28 were female and 41 were male (no one chose

the *other* category). Most were sixteen years of age, with a few still being fifteen. The majority reported having lived their entire lives in Norway (65 students). One student had lived a year in the US, another had lived ten years in Sweden, one reported having been born and lived the first seven years of his life in the Netherlands, and one student had lived in Slovenia. All students put Norwegian and English as languages they could read, write, speak, and understand. Quite a few students also put Swedish and Danish as part of their language knowledge, which is unsurprising among Norwegian speakers, as the languages are similar enough to afford mutual comprehension in most cases. The student who had lived in Slovenia knew Croatian and Slovenian. Many of the students put either German, French or Spanish as languages they could partially read, write, speak, or understand. It was revealed that many of these claims were due to these students having a third language as a foreign language subject in school and they admitted that their competence in these languages was often quite limited. These metadata show that there is a relative gender equilibrium, with a slight skew toward male. The students' cultural backgrounds appear somewhat homogenous with a few exceptions if judged from a nation view and the degree of experience from living in other countries. Only one student had lived in a country where English is the first language. There are several language resources in these classrooms beside Norwegian and English, but proficiency was never measured so the extent of these resources are unknown; however, what can be said is that these students are used to engaging with and learning a range of languages outside of their native tongue.

4.3 The First Research Phase

The first phase had an explorative focus where the beliefs, perspectives, preferences, and previous experiences of the students and teachers were investigated. These elements were ultimately analyzed to answer the questions "How familiar are upper secondary school students with corpora?*"* and "What beliefs do teachers express about corpora as a pedagogical tool?" in the dissertation's first article. The research design consisted of an online questionnaire to the students, which was followed by individual research interviews with the teachers. In the following section, the research design will be discussed in depth by looking at the operationalization and implementation of the different methods. Methodological reflections related to these choices will be made along the way.

4.3.1 Student Questionnaire

A 36-item online questionnaire was distributed to the students through a link provided by their teachers and completed in the classroom during lessons (see Appendix 2). As shown in Table 4, the questionnaire was predominantly comprised of Likert scale items, but included open-ended questions, multiple-choice questions, and a yes/no question. Students could select either an English or a Norwegian version; the majority chose Norwegian. The questionnaire followed a predetermined sequence, discussed below, with predominantly conditional questions, i.e. the students had to answer each question in sequence without the option of skipping ahead. An exception was the question inquiring about corpora directly; if they answered negatively on whether or not they had heard about corpora, they would automatically skip the next question concerning *what* they had heard about corpora. The questionnaire was created by the researcher using the online Checkbox tool (Checkbox.com, 2018), which offered a range of ways to customize and present questions and answer categories, a simple data analysis, and data exportation options to Microsoft Excel or statistics programs. The program was connected to the university server for secure data storage. The implications associated with the data types produced through the questionnaire (see Table 4) are discussed in the Section 4.7.

Table 4
The items and data types produced in the questionnaire

Item type	Data type	Item frequency
Likert scale items	Ordinal	24
Open questions	Nominal	3
Multiple-choice	Nominal	8
Yes/no	Nominal	1
Total		**36**

The Likert scale items consisted of statements where the respondents had the options to *strongly disagree, partially disagree, neither agree nor disagree, partially agree,* and *strongly agree* (with the exception of the two items where students evaluated their language skills, as shown in Section 4.2.4). The five-point scale allowed for a central option for the respondents who did not want to take a side, and thus avoided the assumption that they had a clear opinion on the matter. The design choice of *strongly agree* over the sometimes-used *agree* was made because the latter has an air of definitiveness, while the former has a clearer evaluative component

implying a relative strength of agreement greater than the *partial* alternative. The same scale was used for all of these items to make comparisons easier and avoid confusion.

No immediately similar studies were found from which an already tried and tested questionnaire could be used for replication, and it was therefore necessary to construct a questionnaire from the bottom up that fit this study's aim. The questionnaire was created following guidelines by Cohen et al. (2018; below), and the items were based on previous DDL research, and theoretical constructs associated with technology integration under the broad themes of pedagogy, technology, and corpora (see also Article 1).

As one guideline when constructing a questionnaire, Cohen et al. (2018, pp. 489-505) suggest that the order of the questions be carefully considered so as to avoid unnecessary priming effects or negative affective effects on the respondents. Although few of the items were considered by the researcher to be especially sensitive, one cannot rule out that certain items could have touched on sensitive areas for some students. Items deemed to have a negative or positive influence on the students' affective filters, such as questions about their self-perceived English proficiency or their attitudes toward the English subject, were therefore placed toward the end of the questionnaire. Meanwhile, items concerning general reflections about their experiences and habits with digital technology were placed at the start of the questionnaire. Direct questions about corpora were also placed toward the end of the questionnaire so as not to color the questions about digital habits early on.

Respondent fatigue was also considered as a potential methodological concern. The rigid sequence the students had to follow without the option to skip questions could potentially have led to fatigue or annoyance. To alleviate these pressures, the abovementioned five-point scale offered a less committing, central option as a more neutral choice. While opting for an even numbered scale – excluding the central, non-committal option – would push the respondents to make an evaluative decision, the value-neutral center of the scale was chosen for the following reasons: first, as mentioned above, this option can help alleviate some of the pressure of taking a position where the individual may not have one, thus reducing the potential annoyance with the process. Annoyance might well lead respondents to lean toward the disagreement side of the item or just click randomly. Second, from a research-ethical point of view, forcing someone to express an opinion when they might not have one could result in participants feeling misrepresented or coerced. Third, since the study deals with evaluative statements and perspectives, selection of the neutral option is also interesting, as it could suggest disinterest in the subject under question. In other words, lack of positioning also speaks volumes about the

impact of the phenomenon under investigation. In addition, the questionnaire was kept rather short (36 items) to further avoid respondent fatigue.

The questionnaire was piloted by a smaller group of students from a different upper-secondary school. The pilot included an informal conversation with the participants about the difficulty and structure of the questionnaire, whether or not any item in particular stuck out, or whether or not they thought something was missing. One of the intentions of this pilot was to ensure that the questionnaire items were properly adapted to an upper-secondary school audience, concerning both content and language, so as to avoid confusion or frustrations, and to get an indication of how long the questionnaire would take to complete. In addition to the student pilot, several colleagues reviewed the questionnaire before it was distributed to ensure functionality and to comment on language choices and errors. These reviews led to minor revisions.

Lastly, since the questionnaire was completed during class, it likely increased the number of participants who completed it and eliminated issues such as participants interfering in each other's processes or using the internet to find answers to some of the open questions and multiple-choice questions. For instance, one question asked students to define corpora, which could have been achieved by means of a simple internet search. However, there were no signs of this happening. A potential drawback of filling in the questionnaire in class was that the familiar class environment could have made students feel compelled to participate or pushed them to answer in accordance with how they might think their teacher wanted them to answer. The only step taken to counteract this directly was to emphasize the voluntary nature of the project and the students' anonymity in the information letter (cf. Section 5.8).

4.3.2 Teacher Interviews

The four teachers were interviewed following a preliminary analysis of the questionnaires. The interviews were done one-on-one between the teacher and the researcher at the respective schools of the teachers. The teachers were interviewed in Norwegian. The interviews were then transcribed, and finally translated into English for the publications. The interviews were audio-recorded, which the teachers were informed about and consented to beforehand in writing. Each interview was planned to last about an hour; however, Marcus's interview came closer to two hours due to his longer elaborations, while Nora's, John's, and Sarah's were just past the one-

hour mark. The interviews were semi-structured, which meant that an interview guide was followed (see Appendix 1), but informant supplementations and digressions were encouraged. According to Brenner (2006), "a semistructured protocol has the advantage of asking all informants the same core questions with the freedom to ask follow-up questions that build on the responses received" (p. 362). The interview guide had the same broad themes as the questionnaire, namely pedagogy, digitalization, and corpora, with a range of questions under each of these categories. The informants were given ample opportunity to digress or elaborate, and follow-up questions related to their elaborations/digressions were asked until they appeared to have exhausted the topic. The interview guide was then consulted again to bring the interview "back on track".

Brenner (2006) says to "take advantage of the format" in open-ended interviews "by asking informants *how* and *what* questions that cue informants to give their perspective in their own words" (p. 363). Similarly, Brinkmann and Kvale (2015) suggest prioritizing *what* and *how* questions when conducting the interview, as too many *why* questions can lead to "an overreflected intellectualized interview" (p. 159). These guidelines were largely followed in the interview guides, which predominantly contained *how* or *what* questions (cf. Appendix 1). I made active use of *clarification probes* (Brenner, 2006, p. 364) throughout the interviews by rephrasing the informants' statements as questions so they could confirm, correct, or elaborate on their answers. The questions were initially about their motivations and experiences as English teachers in general, which got the informants talking and helped paint a picture of their general practice. The interview then moved toward digitalization and its effects on educational practice, before directly inquiring about corpora and corpus linguistics in their teaching experience. Finally, the teachers were asked to comment on some of the findings from the student questionnaires. This structure shows the funnel shape typical of open-ended interviews, "beginning with large questions working down to details" (Brenner, 2006, p. 362). The intention of the interviews was to get an impression of the teachers' educational practices, their beliefs about their students and the school environment, and the influence of corpora on their practice. The teachers were informed about the purpose and focus of the research prior to the interviews to promote ethical conduct and transparency. On the one hand, this information could rob the interviewer of more spontaneous responses about corpora. On the other hand, it can lead to more thought-out answers, as the teachers would likely have reflected on the topic before entering into the interview context.

It should be noted that lesson observations could have offered an alternative method to interviewing in order to paint a picture of the teachers' classroom conduct. Although interviews have the advantage of inquiring about someone's general practice, the data produced only point to what teachers *say* they do and not necessarily what they *actually* do. This is not to suggest that they are being dishonest, but rather that human beings are not always aware of the details of their own conduct. This fact should be kept in mind when reporting, interpreting, and reading the results of this – or any – interview study. Additionally, observations could have given supplementary data from the researcher's perspective about how the teachers conduct their practice. Nevertheless, I opted for interviews as these demanded less time from the teachers, and because the sought-after knowledge went beyond general practice. The intent of the study was to access their reflections, judgements, and beliefs, which is not as easily achieved through observations.

4.4 The Second Research Phase

The second phase took place approximately one year after the first phase of data collection was completed. This period involved a two-week cooperation with John and two of his new first-year classes. The students utilized corpus-based tools to solve tasks and write texts about issues in English-speaking countries (see Appendices 4 & 5). Both prior to and during the implementation period itself, the teacher and researcher communicated frequently through meetings and emails. Both the timeframe and topics of this period were set by the teacher in his semester plan prior to the involvement of the researcher, but the researcher suggested a specific period as suitable for corpus resources based on the corpora available (see Section 4.5.2). The research design centered on this two-week period as a case study, and observations, teacher-researcher communication logs, and group interviews were used to collect data. In the following, these methods will be presented alongside their methodological implications.

4.4.1 Planning and Implementation

The implementation period had a case study design in which corpus-based resources from the *Backbone* website (see Section 4.4.2) were introduced. Lesson plans were developed by the researcher with feedback and pointers from the teacher and based on the teacher's original outline of topics and curriculum aims. Although the overall approach can be reminiscent of

design experiments or intervention studies, a softer, less invasive approach was chosen. Instead of treating the implementation of corpora as a *treatment* or *intervention*, which both carry strong connotations that imply there is something that needs to be "fixed" or "cured", the focus was on how, through dialogue and collaboration, the teacher and researcher could cooperate to present helpful resources to the students. We discussed the semester plan set by John, and the researcher suggested a period of lessons set to focus on 'English-speaking countries', 'varieties of English', and 'critical use of digital resources' as an appropriate time to implement the corpus resources.

The two-week period amounted to ten lessons per class. The first two lessons followed shorter discussion and exploration tasks to get the students familiar with the corpora. The main bulk of the period was a more open-ended project using the corpora and other resources for creating texts. The period concluded with a lesson where they discussed their experiences and findings while working with the different digital resources. The final lesson also entailed an evaluation of the applied tools and resources (see Section 4.4.3 for lesson plan details). The lessons sought to facilitate student-active, collaborative learning by mainly using group work and tasks that required discussion and presentation of results, as well as dialogue about the usefulness of the tool. These aspects of the classroom organization and tasks were meant to be conducive to collaborative dialogue and metatalk (Flowerdew, 2015) as well as peer-mediated learning (O'Keeffe, 2020; cf. Section 3.2.2).

John had already chosen five learning aims from the curriculum that the corpus-integration period was based on. Two aims were about the critical and independent evaluation of technology and information sources. Specifically, under the curriculum heading *language learning*, the aim was "[to] evaluate different digital resources and other aids critically and independently, and use them in own language learning", and under *written communication,* the aim was, "In reading and gathering material, the pupil needs to evaluate the content from sources in an independent, critical and verifiable way". In line with these aims, the students were tasked with using and evaluating different resources while exploring the contents of the corpora, which means the period had an evaluative meta-perspective as well. The other three aims concerned English-speaking countries and varieties of English, and therefore had a greater content and language focus. Under *Culture, society and literature,* the students were supposed to "Discuss and elaborate on culture and social conditions in two self-chosen English-speaking countries", under *Oral communication*, they are to "listen to and use native varieties of English from the chosen countries", and finally under *written communication*: "use conventions for

English language construction in order to communicate effectively in writing." The latter three of these five aims had already been worked on through previous lessons but carried over into the implementation period. John had already worked on some of these aims before, which meant that they could be considered partially covered.

This design can be criticized for the relatively short time frame afforded to it in spite of the volume of planned activities. This criticism is valid in that it would conceivably take a lot more time to familiarize oneself with corpus techniques and achieve some degree of corpus literacy (see Mukherjee, 2006). However, the project is based on examining corpus application in a naturalistic context, not in ideal test conditions, which necessarily entails shorter timelines, limited resources, and competing interests and attentions. The issues of time investment and the challenges associated with novelty of working with language and digital corpora are discussed at length in Chapter 6.

4.4.2 Corpus Materials

The *Backbone* [BB] webpage was the primary corpus resource used for implementation in phase 2. It hosts two other multimedia corpora that were utilized during the case study. These two other corpus projects, *The English Language Interview Corpus as a Second Language* [ELISA] and *the System-aided compilation and distribution of European Youth Language* [SACODEYL], served as forerunners to *Backbone* (Kohn, Hoffstaedter, & Widmann, 2009), and are now integrated into the webpage, making it a multi-corpus website. These different corpora can be accessed through the *Backbone* webpage through an easy-to-use drop-down menu, where the English language choices are BB English, BB African English, BB English as Lingua Franca, IVY English (formerly the ELISA), and SACODEYL English. All corpora consist of videotaped interviews of native, EFL or English-as-a-lingua franca [ELF] speakers that have been transcribed. The transcriptions have then been annotated based on topics, a selection of lexico-grammatical categories, and discourse features.

These corpora are *pedagogical corpora*, and their use has been championed by several scholars overusing linguistically oriented corpora for pedagogical means (cf. Section 1.2.2). Pérez-Paredes (2020) describes *pedagogical corpora* as topic-driven, as seeking pedagogical representativeness, and as challenging traditional corpus-search behavior (p. 69; cf. Section 1.2.2). Given the high language levels and degree of linguistic annotation of traditional corpora, the pedagogical corpus approach was chosen over utilizing traditional corpora like the BNC

(cf. Section 1.2.1). The *Backbone* webpage was chosen for several reasons. First, its multimedia design offers spoken language interviews in video and audio format, as well as searchable interview transcriptions. These options provide several ways of examining data and are in line with Flowerdew's (2015) argument that multiple access points to data facilitate knowledge construction (see Section 3.2.2). Second, as mentioned above, the website grants free access to similar corpora such as the BB, SACODEYL, and IVY corpora, which do not seem to be accessible elsewhere. Third, the annotation of SACODEYL is done by teachers "who selected sections of the interviews they considered particularly appropriate for language learning in secondary education" (Pérez-Paredes, 2020, p. 75). Thus, the annotation scheme is based on the pedagogic competence of people who engage in the practice field. Fourth, the website is free, online, requires no downloads, registrations, or emailing for permissions, and has no paywall. All of these factors contribute not only to the didactic potential of these corpora, but also the feasibility for teachers with limited time and resources to make use of them. Furthermore, these corpora appear to satisfy Braun's (2007) suggestion for pedagogically-applied corpora to be (1) more coherent, containing similar texts, and (2) complementary to school curricula (pp. 308-309). The extent to which the latter is the case is discussed in Chapter 6.

Nevertheless, there are certain issues connected to pursuing pedagogic corpora as an option. One such concern is that they are rather rare. In fact, *Backbone* appears to be the only website giving access to the abovementioned corpora. I have been unable to locate any other similar resources online. Another multimodal corpus aimed at younger language learners named *the Multi-modal Corpus Tool 1.1,* or MmCT 1.1 (Hirata, 2020) is being constructed; however, this application was not accessible at the time of the current project. The website CLARIN (CLARIN, 2012) was utilized alongside an extensive literature search to localize suitable resources for pedagogical corpus exploitation below tertiary level. It proved difficult to find such corpora, let alone freely available and free-of-charge ones. The general lack of available resources and the challenge it is to find those that are available can be seen as central points of criticism against the feasibility scenario, as teachers would not know about them, might not bother to look, or may struggle to get access. The issue with few existing corpora of this kind is that they only cover certain aspects of the curriculum due to the particular types of texts and annotations available. The availability of more such corpora with a wider range of interview topics may open the door for greater curricular saturation and higher user rates.

A second challenge is related to the corpora's annotation and their sizes. Firstly, each annotation tag may only show up a couple of times or only once across all the texts. This may be due to the relatively short length of each interview, the tendency for interviews to go in different directions thematically, or the idiosyncrasies of each interviewee. In other words, not all categories appear in all texts. This fact makes comparison between texts and the available frequency information challenging at times. For instance, only a small selection of interviews has been annotated for the use of future referencing expressions, which makes the 'abundance of examples' argument for using corpora in the classroom somewhat moot in these cases. Secondly, many of the interviews are quite short, which results in each text only containing one or a few example(s) of a given phenomenon. Frankenberg-Garcia (2014) found that "multiple corpus examples helped both undergraduates and younger, secondary school students" (p. 140) in their language learning. Sketch Engine for Language Learners [SKELL] was therefore introduced as a free and accessible way to find simple frequency information and more language examples, although SKELL retrieves examples from large general corpora. The Longman Dictionary of Contemporary English was also introduced as a resource that provides corpus-based language examples, definitions, and meta-linguistic information. Thirdly, the corpora contain few individual texts, which brings to question what a corpus is and if smaller corpora can satisfy the main arguments for using corpora in education at all. The texts are authentic, insofar as they are communicated by English speakers, albeit from interviews conducted with the purpose of constructing corpora. If one is interested in using corpora because they give frequency information from representative samples of text, then *Backbone* is less suitable. However, by using supplementary resources, these issues may be alleviated. Furthermore, these types of corpora can serve the dual purpose of (1) introducing students and teachers to new ways of working with language, and (2) help move the scholarly discourse away from seeing pedagogical applications of corpora in the narrow sense of solely analyzing concordances and frequency lists of large general corpora. The feasibility scenario and pedagogic corpora are further discussed in Chapter 6.

4.4.3 Lesson Plans and Tasks

The lesson plans were inspired by guiding principles suggested by Braun (2006) in her work with the ELISA corpus. These principles are (1) a warm-up pre-corpus work like quizzes or awareness-raising, (2) detailed work with one text or similar sections across texts focusing on

lexico-grammatical features, topics, or communicative function, and (3) global work with the entire corpus or larger parts of it in explorative tasks and project work (p. 19). She thus outlines an incremental approach from the guided, particular work toward more open-ended, explorative work. The lessons therefore open with a general discussion about language awareness, before moving into concrete exploration of the corpus in a "learn the language, learn the tools" approach using a combined task and corpus guide pamphlet developed by the researcher; this took up approximately three out of ten lessons for each class. These structured tasks were intended as mediating artefacts for object-regulation (Lantolf et al., 2015; see also Section 3.2.2) to help facilitate the students being familiarized with the tool and provoke discussion and discovery. In the remainder of the lessons, the students chose between two open-ended projects utilizing the corpora of *Backbone* more freely. This approach aimed to promote autonomous corpus-aided language learning as the students got more familiar with the available tools and their affordances. The teacher warned of the students' sometimes-low technological prowess beforehand, which made the guided approach seem all the more necessary.

In addition, the decision was made to de-emphasize theoretical knowledge about corpora during the lessons. Frankenberg-Garcia (2014) makes an important point when she writes that "there is no reason why teachers should confuse learners with corpus-linguistics terminology when it is perfectly possible to give instructions using general words like sentence or sentence extract instead of concordance" (p. 5). She further states: "[l]earners should receive very specific guidelines on what to look out for in the concordances or they will probably not understand what the purpose of the exercise is" (ibid.). In practice, this did not mean that we shied away from any mention of *corpora*, but rather that the guide used general terms as much as possible instead of using corpus terminology with the learners.

The topics of the different tasks were based on the teacher's request and on how well they might demonstrate central features of the tool. For instance, the teacher said the students struggled with idioms, so exploring idioms became the focus of one exercise that simultaneously introduced the students to searching for annotated topics across the corpus, highlighting these annotations, and interpreting them through context. It also tasked them with first discussing what idioms were, and then going to the Longman Dictionary of Contemporary English to find a definition and corpus-derived examples. Many of the tasks therefore became a microcosm of the overall process suggested by Braun (2006; above).

4.4.4 Classroom Observation

Each lesson was observed by the researcher, videotaped, and audio recorded. Since the aim of the study was to get an impression of the students and teacher's work with corpora, teacher-student interaction, student-student interaction, and student-corpus interactions, I chose the role of the silent, non-participating observer as far as it felt natural to do so. According to Bjørndal (2013), observations can range from low to high degrees of openness, and low to high degrees of participation (pp. 46-48). Since my primary role was observation, not teaching or guiding, the degree of participation was considered low. Conversely, the degree of openness was considered high, as the information and consent letter clearly outlined my intent and procedure to the teacher and students.

The lessons were videotaped using two cameras on stands from two different points in the classroom. Three audio recorders were also placed in the classroom to ensure better audio quality and to pick up the verbal interactions of the different groups of students. Bjørndal (2013) suggests that if you are not the pedagogue in the situation being recorded, you should minimize how much you disturb the proceedings (p. 87). However, one should not keep up a pretense that one is not there but make introductions to the students and emphasize the confidentiality agreement to the participants (p. 88). I therefore introduced myself to the class and reminded them of why I was there, as well as the duration and purpose of my stay. I also interacted with the students when they chose to make contact, as anything else would have been perceived as artificial behavior on my part.

The amount of recording equipment may seem a bit excessive. In fact, the teacher informed me that the idea of videotaping at all was what had given the students most pause when considering whether or not to participate. It was therefore important to emphasize that only I would be reviewing the recordings, in order to soothe some of their fears. The choice to use all this equipment was made for a fear of losing valuable interactions in the chaos of large classes. The classrooms were physically quite small, barely fitting the equipment and the researcher, which made the sound chaos intense among more than 30 students. The amount of recording equipment was thus deemed necessary in order to pick up conversations and distinguish between particular speakers.

4.4.5 Group Interviews

Four groups of five students participated in group interviews (see Appendix 3). Interviewing all 72 students was a too extensive process, so a selection was made based on a preliminary analysis of the observational data and teacher recommendations. In all 20 out of 72 students were interviewed. Opting for group interviews allowed for thick descriptions, vivid discussions, and deeper probes of student experiences. According to Brinkmann and Kvale (2015), group interviews "can bring out lively interpersonal dynamics and show the social interactions leading to the interview statements" (p. 333). Group interviews do not seek consensus or conclusions, but rather to encourage a plurality of views on a given topic (p. 175). Instead of individual students expressing their opinion in a one-to-one interview, the group interview can give insight into how groups of students negotiate meaning between peers, co-construct knowledge, and mediate argumentation. Since the interviews sought to uncover a wide array of opinions, possibilities, and challenges tied to direct pedagogical corpus use based on shared experiences and perceptions, as opposed to acquiring expert knowledge, the dialogical format of group interviews was more desirable and suitable.

In addition, group interviews had the added advantages of reaching more of the students than individual interviews would have, while still opening up for rich descriptions and opportunities to follow up on interesting arguments. Moreover, since students are in the presence of their peers, group interviews may help disrupt the strong power dynamics between researcher and informants. For one, the interviewer takes on more of a moderator role, which can entail less authoritative engagement with the interviewed subject. What is more, the group dynamic can promote student discussions as opposed to student-research question-answer interviews. This can ideally lead to knowledge creation on horizontal dimensions of age and experience, as opposed to a strongly hierarchical one. Moreover, if the constellations of learners are made with sensitivity to student relationships, it can facilitate a feeling of safety among the participants. This being said, the presence of the researcher as mediator and questioner, and the presence of recording devices, all contribute to making an artificial context where there are still skewed power relations. For instance, the students know why I am there, and that I have contact with their teacher, which might make them inclined toward positive responses. It is therefore important to stress their relative anonymity, and to minimize my role in the discussion.

There are further potential pitfalls with group interviews. One such pitfall is that the transcripts of these kinds of interviews can become chaotic due to the chaotic nature of group dialogue

(Brinkmann & Kvale, 2015). The potential noise is difficult to plan for, but certain steps were taken to avoid loss of data. One was to utilize two different sound recorders at slightly different positions in the interview room so that each recorder had a chance of picking up half of the group more clearly. Another step was related to the craftsmanship of the interviewer and the importance of listening closely and being vigilant so that clarification probes could be asked after unclear or messy dialogue. A second pitfall was related to group constellations. In any group, one runs the risk of certain voices being dominant, while others are consequently marginalized, or that the dominant opinion in the group drowns out the less represented opinion. Here, gentle probes from the interviewer directed at specific participants or opinions were one way of counteracting this issue. Also related to the domination of a certain opinion, probes were the interviewer temporarily abandoned the quest for neutrality to play the devil's advocate and voice less popular opinions could be useful to bring nuance to the discussion beyond simple agreement.

4.5 Analyzing Data

The analytic approach of both research phases were a process of *segmenting, coding,* and *reassembling* (Boeije, 2010). This approach entails fragmenting the raw data into categories, which are labeled with a code, before reassembling them into new coherent wholes through the lens of theoretical constructs and/or emergent concepts. In the first phase, both the questionnaire and interview data were integrated in the reassembly process to compare teacher and student utterances by using similar codes for the segmented categories. Likewise, during the second phase, the data from teacher-researcher conversations, observations and interviews were segmented into categories and coded with similar code tags, before they were reassembled across data collection methods for comparison. This can be seen as a form of *triangulation* through *complementarity* where convergence is not the goal, but different research methods examine different aspects of the phenomenon and the "[…] separate components are then fitted together like a jigsaw puzzle" (Smith, 2006, p. 465). In order to fragment and then reassemble the data, coding schemes were developed for each of the phases. Prior to data collection, a partial code was made based on broad themes in the literature. Following the data collection, these broad themes were segmented and expanded on based on goodness of fit with the data material, i.e. there was an underlying theoretical foundation, but the codes were largely emergent. According to Boeije (2010), a literature review "[…] may result in the formulation

of a skeletal framework that guides the research process" and that this knowledge "[…] heightens the theoretical sensitivity of the researcher" (p. 89). Brinkmann and Kvale (2015) make the distinction between *concept-driven coding*, where the codes are determined prior to implementation, and *data-driven coding*, where the codes emerge while working with the data material (p. 228). In this study, the back and forth between theory and data both prior to and during the analytical processes can therefore be seen as *abduction*, as opposed to a purely deductive or inductive approach. The strength of such an approach is the possibility for the researcher to learn from and adapt with the data, open up to conclusions beyond preset theoretical ideas, and simultaneously place the novel results in a larger theoretical context.

The final coding scheme of the first phase had the following elements:

- Background information
- Teaching and learning strategies
- Perceptions about teaching
- Perceptions about learning
- Teaching and learning preferences
- Familiarity with digital tools
- Digital proficiency
- Digital preferences
- Familiarity with corpora
- The influence of corpora

In addition to these categories, certain open questions from the questionnaire were also coded and presented in graphs. These categories emerged from the student answers without any premade codes set beforehand. In the list below, the open-ended questions are presented as filled bullet points with the different codes listed as empty bullet points.

- What do you use digital tools for the most in English class?
    - Find information, sources or facts
    - Do tasks
    - Check spelling, grammar or pronunciation
    - Translations
    - Learning games (Kahoot, Quizlet)
    - Write/take notes
    - Watch videos
    - Read texts
    - Learning platforms (Canvas, Itslearning)

    - Presentations/projects
    - Familiarize myself with a topic

- Which digital tools and/or webpages do you use at school?
    - News outlets
    - Streaming services
    - Online lexicons/wikis
    - Online dictionaries
    - Search engines
    - Learner platforms (Canvas, Itslearning)
    - Hardware
    - Microsoft Office
    - Sites for learning games and resources
    - Non-specified webpages/the internet
    - Specific webpages
    - Don't know

The codes of the second research phase reflect the focus on several aspects of the current research, such as pedagogic and didactic choices, student and teacher perspectives, digital and technological experiences, and impressions corpus-specific topics. The second phase utilized the following codes:

- Audio-visual material impressions [AVM]
- Classroom organization and student autonomy [COA]
- Comments on working with corpora [CWC]
- Comments on digital technology work [DTW]
- Distractions, disruptions and digressions [DDD]
- Experiences of the integration period [EIP]
- Issues related to tasks [IRT]
- Language barriers and challenges [LBC]
- Language-learning opportunities including metalinguistic talk/language awareness [LLO]
- Other learning opportunities (beyond linguistic) [OLO]
- Peer scaffolding and interactions [PSI]
- Social and cultural topics [SCT]
- Student's motivation and shaming of other students [MOS]
- Student's previous experiences [SPE]
- Teacher-student interaction/communication [TSI]
- Website design impressions [WDI]

As Boeije (2010) points out, this process of segmenting, coding, and reassembling is not just technical, but entails researcher interpretations, meaning-making, and judgement of fit or misfit

of data points in a given category (p. 77). It is therefore a subjective approach prone to researcher biases and misinterpretations. One measure taken to increase transparency was through openness in the presentation of results. In the case of the questionnaire data, this was achieved by largely presenting percentages and frequencies and discussing them in that form. When the interview data were concerned, openness was sought by presenting the teachers' own words as much as possible to avoid too much unfortunate paraphrasing, although some paraphrasing was necessary to give a broader range of answers space within the limited scope of a research article. Another measure taken to ensure transparency in the interpretive process was a priority of focus on multiple possible interpretations of the data in the discussion. This emphasizes the multiplex and sometimes conflicting nature of qualitative inquiry and knowledge construction, and displays the interpretive process involved in qualitative data analyses, as opposed to reporting "truths". As Cohen et al. (2018) point out, analysis of interview data "is less a completely accurate representation (as in the numerical, positivist tradition) and more a reflexive, reactive interaction between the researcher and the decontextualized data that are already interpretations of a social encounter" (p. 524).

The types of data produced in the questionnaire (see Table 3) also set certain restrictions on the analysis. Nominal and ordinal data are typically considered non-parametric, as there is no way to determine the exact intervals between the items' answer categories (Cohen et al., 2018, pp. 726-727). Since many of the statistical tests assume parametric data (see Cohen et al., 2018, p. 842 for an overview), this assumption, alongside the non-probabilistic sample of this study, made sophisticated statistical testing problematic. Another reason for the somewhat simplified presentation of numeric data was the transparency afforded by frequencies and percentages, which made it easier to analyze the questionnaire data with the corresponding categories of the interview data. Perhaps some interesting results could have had emerged if a correlation test – e.g. Spearman's rho, which allows for correlation testing of non-parametric data – had been used; however, I opted not to test for correlation, so this remains an unrealized opportunity.

Having described and reflected on the data-collection methods of each phase, we now move to the research's transferability and credibility.

4.6 Transferability and Research Credibility

4.6.1 Transferability and Replicability

In both phases of the project, *transferability* replaced the concept of *generalizability*. The focus on socially and historically bounded contexts is arguably a strength of the social-science endeavor. This contextuality provides a situatedness that makes abstraction to a general population problematic. However, this does not have to mean that the findings in this study are relegated solely to the local context. Schreier (2018) explains that the key notion of transferability is not "[…] to generalize to an abstract decontextualized population, but to determine whether the findings obtained for one instance or set of instances in one specific context also apply to other instances in a different context" (p. 86). When the term *generalization* is applied here, what I mean is *statistical generalization*, which is the gold standard in quantitative research for obtaining *external validity*, i.e. whether the results can be mathematically extrapolated to a larger population represented by the sample. Alternatively, Brinkmann and Kvale (2015) describe the term *analytical generalization* as "reasoned judgment about the extent to which the findings of one study can be used as a guide to what might occur in another situation" (p. 297). Similar to descriptions of transferability by Schreier (2018), they emphasize the need for rich – or *thick* – description, shared reader and writer responsibility, and the judgment of fittingness of one situation with another based on context comparison (Brinkmann & Kvale, 2015, p. 297). In this study, thick descriptions were provided through description of the institution of upper-secondary school in Norway, the participants' backgrounds, the classroom environment observed, the researcher's position and participation, and by presenting methods of data collection, extracts from the transcriptions, tasks and lesson plans, and multiple interpretations in a transparent way. The shared reader and writer responsibility is tied to context comparison in that the writer provides as thick, transparent descriptions as possible, while the readers judge the fittingness to other situations. The role of reader was also taken in this study through the literature review and the comparison with previous reported results to the ones in this one. Moreover, the lesson plans, the BB tasks and guideline pamphlets, the interview guides, and the questionnaire are all included in the appendices of the dissertation (see Appendices 1-5). Together with the thick descriptions of the research context, these allow for transparency regarding the research process and open the door for replication studies to be carried out. Arguably, replicating all or parts of the research in other

contexts and with other participants would strengthen the results of this study through multi-context comparisons.

Although the questionnaire in phase 1 can be said to produce quantitative data, I elected to avoid any statistical generalization tests. The reason for this is two-fold. First, the sample of students was relatively small (n=154) compared to the population of upper-secondary school students during the 2018-2019 school year, which was 188,482 (The Norwegian Directorate for Education and Training, 2019). Second, since purposive sampling seeks to "[…] select instances that are *information rich*" (Schreier, 2018, p. 88), it is said to be "[…] deliberately and unashamedly selective and biased" (Cohen et al., 2018, p. 219). Conversely, statistical generalizability operates under the assumption of probabilistic sampling through randomly selected participants. Thus, the purposive approach outlined in Section 4.2 operates under different assumptions and is consequently inappropriate if statistical generalization is the goal. Furthermore, the many open-ended and multiple-choice items resulted in nominal data, which were not always quantified, but which proved a central part of the reported data in the first article (see Article 1).

4.6.2 Trustworthiness

Both the questionnaire and the interview guide of phase 1 were piloted. The questionnaire was piloted through a smaller group of students who first completed it individually, then engaged in a group conversation with the researcher voicing their impressions, understanding of certain questions, and interpretative issues. In addition, before distributing the questionnaire, several colleagues, both linguists and teacher educators, reviewed and commented on content and language of the questionnaire draft. In phase 2, the implementation and group interviews were not piloted, but the task and guideline pamphlet made for the *Backbone* website was sent to two English teachers to be tested and commented on. The design of this pamphlet and the lesson plans were also discussed with the teacher, John, prior to implementation. Paraphrasing of informant utterances for clarification and probing questions were utilized in both the teacher and student interviews of both phases to confirm the researcher's interpretation of the teachers and students' answers.

4.7 Ethical Considerations

4.7.1 Informed Consent and Confidentiality

Several steps were taken to ensure ethical conduct. First of all, a detailed description of the research, including the group of participants, the type of data sought after and how the data would be handled and stored, were sent to the Norwegian Center for Research Data (abbreviated NSD in Norwegian), owned by the Norwegian Ministry of Education and Research (nsd.no, 2019), for review and approval. In addition to a detailed description of the research process and data handling, the NSD also requested to see the information letters and consent forms going out to all participants. These were all approved before data collection began (see Appendices 6 & 7).

Written consent was requested from all participants, and those who elected not to sign consent forms were excluded from the research. All participants were given a written information letter outlining in detail the research process, their role in it, and what type of information they would have to supply. It also informed them that participation was voluntary, that they could refuse to participate, and that they could withdraw their consent at any time during the research process without fear of consequences. This letter was therefore in line with the Norwegian National Research Ethics Committees' (NESH) guidelines for consent, which state that consent should be given freely, expressly, and informed, and that it should be documented (NESH, pp. 14-15). Since all participants were fifteen years of age or older, it was sufficient to request consent from the participants themselves, i.e., parental consent was not required, as minors who have turned fifteen can usually consent to research collecting and using their personal information (NESH, p. 20). The consent form and information letter were distributed on paper by the teachers for their students to fill out, and signed copies were returned to the researcher by the teacher in person. Thus, signed consent for all participants has been documented and stored in physical form.

Anonymity was ensured in several ways. The teachers were given pseudonyms so that only the researcher knew their real names, and the connection between the teacher, the school at which s/he worked, and their students were made through a coding system only familiar to and accessed by the researcher. The codes were solely used to connect teacher with a group of students, not to target individual students. When the teachers were asked to comment on their students' answers, those answers were shown as diagrams of either all respondents of the study

or all respondents from their classes. Only the counties where the schools were situated were revealed, as well as the general direction of the classes' education (e.g. 'general studies' or 'vocational'); these areas of geography and education were considered to have large enough teacher and student populations to make identification of specific individuals or institutions highly improbable. The students were not required to give their names in the questionnaire. The interviews did not require the teachers to identify or discuss specific students, and if students mentioned a teacher in the questionnaire, the names were replaced by pseudonyms during transcriptions. All data and codes were kept on a university-provided computer in a locked office, or on the questionnaire's database stored on safe university servers.

4.7.2 Other Ethical Concerns

There are some potential ethical issues inherent in this type of study related to the intersection between teacher practice and beliefs, student learning, and the research community. In the case of data-driven learning and corpus linguistics, there is a vocal community of corpus linguistics who, by seeing or theorizing about the benefits of exploiting corpora pedagogically, promote corpora's place in the classroom or even their transformational potential on pedagogy, as well as lamenting the lack of teacher interest in the subject. One must be careful of such advocacy, since it can lead to biases "distorting research to favor one's cause or conviction" (Strike, 2006, p. 58). Not to suggest that corpus linguists with a pedagogical aim are to be considered an interest or advocacy group with surreptitious or nefarious intentions, but rather that even small biases must be accounted for and discussed in a transparent manner. Our intentions should not be to "sell in" corpora to the teachers and students, but instead to see *if* and *how* corpora fit their everyday practice.

There are several ways of tackling the advocacy issue. First, it is up to the researcher to offer transparency in the written product to make his position clear to the readers. Second, it should be noted that the PhD project is funded by Inland Norway University of Applied Sciences based on a project descriptions written by the researcher, and that it has not received external funding, so in that sense the research is "free" from policy makers or particular interest/advocacy groups. Third, the researcher must be prepared to criticize established ideas or wishes in the corpus community when discussing the findings. As Strike (2006) points out, "researchers should participate in advocacy research only when they can assure themselves that they can maintain their objectivity, and they should not participate in advocacy efforts where secrecy is expected

or where results will be made public only if they support preconceived views" (p. 67). Although the corpus community might not be defined as an advocacy group in a socio-political sense per se, these guidelines by Strike (2006) remain relevant, the discussion of whether or not 'objectivity' is at all possible notwithstanding. The most important factor in this study, in the author's opinion, is to maintain integrity to report results that might go against the preconceived views of the larger research community in question. These reflections also mark a strength of this project, since the researcher originally has a teacher background, not a corpus linguistics one, and can therefore offer a different perspective. Fourth, when analyzing, discussing and presenting the results, one should avoid the assumption that one particular group 'owns the truth'. For instance, if the teachers' views do not coincide with the corpus milieu's views, one should not assume that the teachers are in the wrong. The same problem may also be found in the discrepancies between student and teacher views, or between a majority student opinion and a minority one. By attempting to report and describe the multifaceted, multiplex range of results, as opposed to favoring one group or the majority, one can avoid unwarranted conclusions while supplying much needed nuance to the issue at hand. Finally, one should be careful not to enter the classroom as a researcher with a mind to transform pedagogy. Although a goal of the research is to add to the repertoire of the teachers and students, this study aims to introduce corpora in a way that is less invasive and more complementary to routines that are already established, and that is sensitive to teacher and student needs and voices.

4.8 Chapter Summary

This chapter has outlined and discussed the participants and the two phases of data collection of this project. The participants and sampling strategy were first described, followed by the research design and modes of analysis of each phase. Lastly, the credibility, transferability, and ethical reflections of the study were covered. Another intention of this chapter was to show the relationship of the different applied methods, the connection between the articles and the overall project, and consequently the integrated whole of the entire study. In the following chapter, the articles that resulted from the data collection are briefly summarized before their connection and overarching themes are discussed in Chapter 6.

5. Summary of the Dissertation's Articles

5.1 Chapter Introduction

This chapter presents a brief overview of the dissertation's three articles (see Articles 1-3) and summarizes their main findings. The articles investigated the direct application of corpora in the upper-secondary classroom with a focus on both teachers and students. Each of them contributed to illuminating the phenomenon under investigation in different ways and to answering the dissertation's overarching research question *How can corpus-based approaches be integrated into Norwegian upper secondary schools and how are they received by the users?* (see Section 1.4.2). The first article (Article 1; Section 5.2) was written based on data from individual teacher interviews in the first phase of data collection, while the second and third articles (Articles 2 & 3; Section 5.3 & 5.4) were written based on data from a classroom implementation of a corpus-based approach to language learning, the observations from this implementation period, and subsequent student group interviews following the implementation period (see Sections 4.3 & 4.4 for a detailed overview of the research phases and data collection). The first article (Section 5.2) is published in the *Nordic Journal of English Studies*, the second article (Section 5.3) is not published at the time of writing, and the third article (Section 5.4) is published in the *Nordic Journal of Language Teaching and Learning*. Both journals are peer reviewed.

In the first article, the focus was on teachers who had had formal corpus training as part of their teacher education program and on the corpus literacy of their students. This work revealed several obstacles to corpus-based approaches in pre-tertiary education. Taking these obstacles into account, the second article reported on a collaboration with one of the aforementioned teachers to integrate a corpus-based approach in two of his upper secondary classes. This study identified several instances of learning opportunities but also exposed challenging factors related to corpus implementations in pre-tertiary education that greatly complicated the process and success of the approach. Building on the experiences of both articles 1 and 2, the third article investigated more closely the issue of teacher and student roles during the implementation period and proposed a new theoretical framework through inquiry (cf. Section 3.3.4) that represents a new way forward for DDL.

The following sections present the articles in the order outlined in the previous paragraph. The three articles are covered in one section each (Sections 5.2, 5.3 & 5.4). The chapter is concluded in Section 5.5.

5.2 Article 1: 'Corpus Literacy and Applications in Norwegian Upper Secondary Schools: Teacher and Learner Perspectives' (Karlsen & Monsen, 2020)

Article 1 reported on the potential pedagogical corpus applications of four corpus-trained teachers and the self-reported corpus literacy of their secondary-school students to answer the questions: "How familiar are upper secondary school students with corpora?" and "What beliefs do teachers express about corpora as a pedagogical tool?" (see Article 1; Karlsen & Monsen, 2020)[6]. It built primarily on two theoretical perspectives, namely *teacher's beliefs* and the *TPACK* [Technology, Pedagogy and Content Knowledge] model. The teacher's beliefs perspectives posit that the many attitudes, ideas, and beliefs of teachers, some of which the teachers are unaware of, impact their practice in different ways, for instances through the ways they view their students or the way they believe learning takes place (i.e., epistemic beliefs). The TPACK model frames the relationship between pedagogy, content knowledge, and technology when (new) technology is integrated into educational practice. This framework was not used to measure this relationship, but to point out descriptively the need to see the implementation of corpora in classrooms as a case of new technology, new pedagogy, and new content. It was operationalized through corresponding categories for both the student questionnaire and the teacher interview guide.

The study showed that the students had little to no knowledge of corpora and the teachers reported having mostly avoided it in their practice with their students. There were only a few exception to these findings, as two teachers had talked about the basics of corpora in some capacity to a few of their students, one of said students managed to give a succinct definition of what a corpus was in his questionnaire response. According to the teachers, they had been deterred from using corpora in their teaching because of the lack of accessibility due to paywalls, registration requirements, and search limits. In addition, the teachers all described corpora in terms of academic research and linguistics, and that the design and use of corpora

[6] This article was co-written with my second supervisor Marte Monsen of Inland Norway University of Applied Sciences. She was the second author and mainly contributed to the theoretical framework and Section 2 of the article.

were too daunting for their students. One of the teachers pointed out that corpora were made completely for academics, and not teachers and students. Another teacher argued that her students probably lacked both interest and competence to properly navigate and exploit a corpus. Of note was the point that the teachers in their own education reported having mainly worked with large general corpora in an academic environment, which likely impacted their impressions of corpora in that direction.

Further reasons were sought in the data to discover other potential barriers to pedagogic corpus implementation. One set of obstacles were categorized as a *pedagogic dimension*. These obstacles were related to the discrepancy between the teachers' largely topic-focused teaching and the linguistically-focused DDL. This discrepancy was linked to the teachers' beliefs about language learning as a mostly implicit process. It was also found that the majority of students regarded their teachers as English knowledge experts, which could be an obstacle to an open-ended DDL where the teacher has a facilitator role and may not have the answer to every question. A final dimension that was investigated was the *digital dimension*. The majority of students considered themselves competent and frequent users of digital tools in an educational context, but the teachers perceived their students as having low digital competence. While the students' positive self-evaluation of their own digital competence indicates that the digital nature of corpora might not overwhelm them, their teachers' negatively skewed beliefs about their students' digital competence may have contributed to their reluctance to introduce corpora as a demanding digital application.

5.3 Article 2: 'Integrating Multimedia Corpora in the Secondary School Classroom in Norway' (Karlsen, in preparation)

Article 2 reported on a study that investigated the direct application of corpora in two upper-secondary classes in Norway in cooperation with one of the EFL teacher who participated in the first study (see Article 2 for the current study; see Section 5.2 & Article 1 for the first study). The study sought to contribute to filling the knowledge gap relating to the use of corpora for learning in pre-tertiary education (e.g., Wicher, 2020) and posed the following two questions: "How can pedagogic corpora be applied in an upper secondary school and how is this experienced and approached by the teacher and learners?" and "What learning opportunities and challenges emerge when introducing corpora directly in the EFL classroom?" This study

built on some of the obstacles unearthed in the first study and used corpus resources designed for pre-tertiary pedagogical purposes (see Pérez-Paredes, 2020).

In the study, one corpus-trained in-service teacher participated alongside two of his classes (student n=69) in a two-week period of corpus implementation. The *Backbone* website was used for this purpose, as it offered one of the few freely accessible *pedagogic corpora* (cf. Pérez-Paredes, 2020; Section 1.2.2). Data collection consisted of teacher-researcher interaction logs, classroom observations during the implementation of corpus-based tasks, and four student group interviews following the implementation (two groups from each class). The focus was mainly on student and teacher voices, student-to-student and student-to-teacher interactions, and engagement with the provided material and the corpus resources.

The study found that distractions and digressions on the part of the students were the norm in both classes. Despite the students having been provided structured tasks that included user manuals of the corpus resource, they often felt lost, and the teacher had to reiterate that they should consult the tasks. Moreover, the teacher expressed that he felt little ownership over the tasks. These observations suggested that assumptions about students' interests, capabilities, and autonomy are overly romanticized.

There were instances of potential language learning through students discussing their corpus searches by using metalanguage and references to their mother tongue. These instances included the discussion of the layout of a frequency list and why certain words were so frequent while others were less so, semantics discussions on the basis of concordance lines, and discussions of idiom searches in the corpus. In addition, there were opportunities for learning about social and cultural aspects of language. The corpora consisted of video clips of speakers of pedagogically neglected Englishes (e.g., Irish English) and the students' engagement with these varieties triggered utterances and discussions that revealed socio-economic prejudices. The teacher used these opportunities to discuss issues of both language and culture with his students.

The student group interviews revealed several obstacles to the integration of corpus technology in upper-secondary school in Norway. Among the issues were the site's design, which was perceived as outdated and messy, the poor video and audio quality, and the subject content in the corpora, which was deemed too general. Further criticism can be directed at factors outside the corpora themselves, such as the short time frame, the teacher's approach, and the task and lesson designs. Despite the proposed advantage of the topic-driven nature of pedagogical

corpora (Pérez-Paredes, 2019, p. 69), a somewhat surprising finding was that several of the interviewed students wanted more language-focused lessons and wanted to de-emphasize topics. Positive student remarks centered on the use of frequency lists, the website's highlighting function, and the text and video multimedia option.

5.4 Article 3: Educational Roles in Corpus-Based Education: From Shift to Diversification (Karlsen, 2021)

Article 3 discussed the educational roles afforded by the use of linguistic corpora as a teaching tool in pre-tertiary education and investigated upper secondary students' opinions and experiences of educational roles following corpus-based lessons (Article 3). The study built on the experiences and interview data from the second research phase (see Section 4.4) but with a distinct focus on roles and responsibilities. The research questions were: "How do DDL proponents' assumptions about the upper secondary classroom and its educational roles align with the experiences and opinions of students?" and "How can perspectives from inquiry-based education and student-centered teaching inform the conceptualization to educational roles in DDL?" The data were viewed through the theoretical lens of *role diversification* in inquiry-based learning and student-centered teaching (see Section 3.3.4) in order to address the obstacles and challenges of the dissertation's second study (Section 5.3; Article 2).

The results show that during the largely student-centered, corpus-based approach, students felt that the teacher was absent and unengaging at times, but that they were used to this sort of "self-study" in his lessons. The students did not describe learner-centered approaches in exclusively positive terms but wanted more explanations and guidance from the teacher. For instance, while writing texts, some of the students felt that the lack of guidance and supervision interfered with their language learning. Some students also expressed that learning the tool got in the way of their English language learning. It became difficult for the teacher to estimate the degree of freedom and responsibility that should be given the students, and too much student responsibility resulted in a feeling of unclear teacher expectations, according to one student. The students also wanted more variation in the way their lessons were conducted in general and not just student-centered groupwork.

The study found that the students expected the teacher to frame the lessons, motivate them, specify the tool's usefulness and areas of use, and aid them in their work even prior to any task-related issues. These expectations show that "role taking becomes a negotiation of involvement, role definitions, and responsibility" (Karlsen, 2021). Thus, it is important to cultivate an environment where both teachers and students have a versatile and diverse toolbox of roles they can apply to different situations, different tasks, and with different learner styles and preferences.

5.5 Chapter Summary

This chapter gave a broad summary of the dissertation's three articles and the links between them. The findings presented above along with the theoretical framework proposed in Chapter 3 are the foundation of the discussion in the proceeding chapter. In the following, these findings will be the cornerstone of the discussion, which in turn will be lifted to discuss a way forward through a theoretical lens and consider the impact of these elements on teachers, students, teacher training, and DDL research.

6. Discussion, Contributions and the Road Ahead

6.1 Chapter Introduction

This chapter discusses the findings reported in the dissertation's three articles (Articles 1-3; see Chapter 5 for a summary) in light of previous DDL research and the proposed theoretical framework of Chapter 3. Articles 1-3 contain discussions of their respective findings, but the aim of the present chapter is to lift the perspective to several core themes that emerged across the articles or from either research phase (see Chapter 4) and to clarify the empirical, methodological, and theoretical contributions of the dissertation to the field of DDL and the practice field. Note that not all findings and discussions are reiterated here (for more in-depth deliberations consult the articles). As discussed in the dissertation's introduction, the use of corpora for educational purposes in secondary school remains "[…] relatively uncharted territory" (Wicher, 2020, p. 31) and there is a prevalent research-practice gap in DDL that teachers and learners in particular have the potential to bridge (Chambers, 2019; Section 1.1). This gap entails that, despite considerable research efforts, the direct applications of corpora are far from normalized in the classroom (Chambers, 2019), and this is particularly true for pre-tertiary education. There are several factors that contribute to the persistence of this gap: few studies examine corpus use in pre-tertiary contexts, as most are set in higher education (see Boulton & Cobb, 2017; Chambers, 2019); most studies focus on the students with courses taught by teacher-researchers, not in-service, secondary-school teachers; little is known about the broader picture of pre-tertiary teachers' decision-making (Wicher, 2020), which may be traced to the general paucity of qualitative studies below university level (Pérez-Paredes, 2020) and the general focus on student-autonomous working methods; and there appears to have been little attention given to the curricular relevance of corpus use (see Braun, 2007), which is particularly true with regard to the relevance of corpus-based approaches for Norwegian learners of English. It was the aim of this dissertation to address several of these research-practice gaps by examining the practice field, utilizing pedagogic corpora (see Pérez-Paredes, 2020) and investigating pre-tertiary corpus applications through collaboration with a teacher. Teacher and student voices were of particular interest because they maneuver the practice field on a daily basis and are impacted the most by changes to practice. The following main research question was posed for the dissertation:

How can corpus-based approaches be integrated into Norwegian secondary schools and how are they received by the users?

The main research question was addressed through two distinct research phases (see Chapter 4) that culminated in three articles. The first phase investigated the corpus applications of in-service teachers who had corpus training, and who were likely candidates to have applied corpora in their practice, as well as examining their students' corpus literacy. When it was discovered early on in the research process that these teachers had, for the most part, avoided corpora in their practice, the focus shifted to why these teachers were reluctant to integrate corpora in their practice by looking at factors relating to them and their students. Several barriers to corpus integration in pre-tertiary education were identified (see Article 1) that had to be circumvented for successful, direct application of corpora. Taking these barriers into account alongside obstacles identified in the literature, the research question was further addressed through a teacher-researcher collaboration in the second research phase, which sought to integrate corpora in the teacher's practice and evaluate this process based on student feedback and classroom observations. Lastly, the main research question was addressed by proposing inquiry-based education as a theoretical framework and a way forward for DDL that meet many of the identified obstacles from the dissertation and the literature.

In addition, this dissertation was set in the intersection between two curricula. Educational developments in Norway have led to the renewal of the subjects and new core and subject curricula, which are being implemented between 2020 and 2023. The theoretical framing of DDL as a mode of inquiry meets several of these developments (see Section 3.3), in particular the call for more discipline-specific and research-emulating methods of learning in schools.

The relevance of the dissertation lies in its empirical, methodological, and theoretical contributions to the practice field (i.e., teachers, students, and the curriculum), to DDL scholars, and to teacher education. Section 6.2 discusses the dissertation's core themes, which include the novelty problem (Section 6.2.1), the relevance problem (Section 6.2.2), and the new way forward with inquiry (Section 6.3). Section 6.4 summarizes the contributions and limitations of the dissertation and proposes some future directions. In Section 6.5, concluding remarks are presented.

6.2 Discussion

6.2.1 The Focus on Teacher and Learner Perspectives

The empirical data that underlie this discussion and form the dissertation's empirical contribution are largely teacher and student perspectives, opinions, and feedback. The dissertation maintains that the methodological trend in DDL to focus on student feedback (cf. Section 1.3) is an important one if DDL is to be successfully integrated into secondary school in a hands-on, direct fashion. However, a central subject of investigation in the dissertation has been the teacher in corpus-based education and how the teacher's role can be reinvigorated in the scholarly discourse. Although the importance of teacher or tutor instructions, scaffolding, and examples were highlighted in several studies (cf. Section 1.3; Crosthwaite & Stell, 2020; Di Vito, 2020; Moon & Oh, 2018), there is a lack of thorough examination of the teacher's role, and the relationships between students and teachers or peer-to-peer remain somewhat nebulous. Meanwhile, there is a prevalent focus on learner perspectives in DDL that could be traced to early conceptualization of DDL as a student-centered approach in which one wishes to "cut out the [teacher as a] middleman" (Johns, 1991, p. pp) and further linked to its associations with constructivism (cf. Section 3.2), which runs the risk of being individualizing with the focus displaced from relationships in education toward the learner, and the concurrent consequence of the vanishing teacher role (cf. Biesta, 2016; Section 3.2.2). In this sense, this research-practice gap has both a *practical-empirical* dimension in that there is little research on the teachers' decision-making (see Wicher, 2020) and roles in the DDL classroom, and a *theoretical* dimension in that DDL has been linked to a student-centered, constructivist principles to the detriment of the teacher role. The dissertation seeks to refocus on the teacher-student relationships and reinvigorate the teacher role in DDL both through its practical application of corpora in the classroom and its theoretical contribution through the perspective of inquiry-based education.

In addition to being mediators and facilitators, teachers are the conduit between corpus linguistics and DDL in education, but they are also gatekeepers of practice. In Norway, teachers have considerable methodological freedom and professional autonomy (cf. Section 2.3.1), which largely provide them the liberty to accept or reject novel approaches to teaching and learning at their professional discretion. Meanwhile, it has been recognized that it is teachers and students themselves who have the potential to close the research-practice gap of DDL most efficiently (Chambers, 2019) as they are at the frontlines of practice.

The first part of the discussion focuses on obstacles that emerged during both phases of the research. Obstacles became the focal point due to two general observations: (1) The teachers in the study, in spite of their corpus training and experience, had avoided the use of corpora directly with their students almost entirely (cf. Article 1), and (2) the students expressed negative opinions about the corpus-based approach and showed a general lack of investment during the lessons (cf. Articles 2-3). Two core themes grew out of the work with the three articles, and these have been termed *the novelty problem* (Section 6.2.2) and *the relevance problem* (Section 6.2.3). These themes represent the primary empirical contributions of the dissertation. In order to address these problems, DDL as a mode of inquiry (see Section 3.3) is proposed and will be discussed in relation to the aforementioned core themes (see Section 6.3).

6.2.2 The Novelty Problem

There is a *novelty space* involved when students are taking on educational approaches that are new (Fiskum, Myhre, et al., 2018; see also Section 3.3.3). This space comprises the distance between working methods the students are familiar with and new, unfamiliar ones. Not only would DDL entail a new way of teaching and learning related to research-emulating working methods such as formulations of hypotheses and research questions, data analyses and interpretation, and inductive learning (see Section 3.3.2), but it would be through a new digital medium and new types of data. The TPACK [Technology, pedagogy, and content knowledge] framework has been used to draw attention to the intersection of pedagogy, technology, and content as factors involved when introducing new technology to the learning process (Karlsen & Monsen, 2020; Meunier, 2020). These factors combined can make traversing the novelty space seem like a daunting task for both teachers and students, as teachers are the ones who need to guide their students in this new environment. Thus, one of the core themes that emerged from the articles was the participating students and teachers' encounter with the novelty space in what I have termed *the novelty problem*. Although student feedback from previous studies has been largely positive (cf. Section 1.3), obstacles that can be related to the novelty problem have been observed. These obstacles included students' insufficient strategies to interpret concordance lines and wordlists (Braun, 2007; Forti, 2020), students' reliance on more traditional, explicit learning approaches and their reluctance to embrace new ones (Moon & Oh, 2018; Szudarski, 2020), and issues not related directly to the corpus approach such as the novelty of the topics covered while working the corpus (Di Vito, 2020). These are all

contributing factors to the novelty space that students may experience, and that the teacher needs to address, to ensure successful corpus integration. Additional aspects of the novelty problem emerged through the dissertation's studies that contribute to our knowledge of the novelty space in a more holistic manner.

Previous studies found that first contact with corpora for younger learners can sometimes be awkward (Papaioannou et al., 2020) and certain students from both secondary and tertiary education find corpus consultation challenging (Chambers, 2007; Szudarski, 2020; see also Section 1.3). In the dissertation's second article, which also reported on secondary-school students' first contact with corpora, the interviewed students expressed that they had felt it more akin to a computer course than an English lesson. Some students experienced that learning to use the corpus tool took away from their language learning due to the time required to learn its functions, and none of the students who participated in the interviews felt like they had learned new English, in spite of the observed learning opportunities. In addition, these issues associated with learning a new tool were further exemplified through the students' criticism of the *Backbone* interface design (see Article 2), which showed that tool complexity, layout, and aesthetics impacted their experience of it negatively. Thus, the training required to utilize the concordancer was experienced as a barrier between the students and their language learning, which stood in the way of exam preparations and more valuable language study. The extent to which the students perceive the tool as relevant naturally impacts their willingness to traverse this novelty space (see Sections 6.2.3 & 6.3.2).

The issues outlined above are largely tool-specific and tied to the computerized nature of corpora. The use of any corpus application would entail a degree of novelty, but the particular type of tool comes with its own novelty. As stated in the introduction (Section 1.2.2), online concordancers should be designed with a view to their simplicity and their functionality (Lee et al., 2015); however, these two factors are bound to impact one another. The many functions, buttons, and option trees of the *Backbone* website gave it a wide range of functions but made it unintuitive, an issue which was exacerbated by some difficult button names (e.g., *determiners and quantifiers*) and dwindling student interest related to their criticism of the website's aesthetics (see Article 2). Although multiple functionalities are one of the arguments for how corpora facilitate knowledge construction (cf., Flowerdew, 2015; Section 3.2.2), it comes with the cost of time, training, and investment. Meanwhile, a more simplistic design with fewer functions may alleviate the novelty problem somewhat, but at the detriment of more opportunities to examine the corpus data. Boulton (2010) suggested that the use of concordance

lines printouts was one way in which the issue of concordancer complexity could be avoided and learners could get direct access to the data. This solution circumvents one aspect of the novelty problem in order to address another – students' struggles with interpreting concordance lines and deducing rules (Braun, 2007; Moon & Oh, 2018; see also Section 1.3; Szudarski, 2020) – but it is not really a viable option for a multimedia corpus like *Backbone* whose main beneficial feature was the intersection between multiple corpus functionalities and video files. Nevertheless, paper-based corpus work is one way in which one aspect of the novelty space can be worked on at a time.

Previous studies have focused on the issues related to learner-to-corpus interactions and corpus design (cf. Section 1.3); however, this focus only provides one piece of the puzzle that is the novelty problem. For instance, the students of the second research phase reported having had varying degrees of experiences from lower-secondary school when it came to writing academic texts or creating research questions, and that they would have preferred more explicit guidance on how to write these types of texts. There are therefore varying degrees of newness involved in learning to write academic texts that depend on the students' previous education, such as hypothesis forming, the use of sources and appropriate citation techniques, or learning to reason and build arguments based on critical thinking and sources. These skills alongside the basic skill of writing are founded in the competence aims in the curriculum for upper secondary school in Norway (Utdanningsdirektoratet, 2013) and the mastery of these skills is pivotal to the success of the students' exam and future endeavors in higher education. The point is that students experience many novelty *spaces* that influence each other and occupy a finite time frame. Consequently, the rather extensive novelty space identified relating to corpus tools and techniques can therefore be experienced as frustrating and as a time thief by students who are already confronted with significant novelty in curricular requirements. While DDL techniques have been found to be useful in text production for both university students (Charles, 2014) and primary school students (Crosthwaite & Stell, 2020), the combined novelty of these different aspects of the English subject and DDL become an obstacle instead of an aid. Arguably, the investment into a new approach will rely upon the relevance of the approach (see Sections 6.2.3 & 6.3.2), the utility and intuitiveness of the tool, and the knowledge and guidance of the teacher.

There is another aspect to the novelty problem beyond students' immediate experience that greatly impacts the dissemination of corpora to pre-tertiary education, and that has not featured in previous studies. As shown in the first article, the interviewed teachers held certain beliefs about their students' digital and language proficiency that, coupled with their knowledge of

corpus linguistics, made them reluctant to introduce corpus work in their lessons (see Article 1). Thus, obstacles arose in the interface between the teachers' beliefs about corpora and the teachers' beliefs about their students. At the heart of the issue is the teachers' ideas about what corpora and corpus linguistics are. The teachers expressed having mainly worked with large, general corpora such as COCA and the BNC in an academic context (cf. Article 1). This limited experience, in turn, influenced their beliefs about corpora as an academic phenomenon and as involving demanding research processes. It is unsurprising that the types of corpora student-teachers encounter and the way in which they have worked with them during their educational journey are formative experiences for their corpus knowledge and that this colors the lenses through which they see the corpora's utility, which in turn translates to a reluctance to overwhelm their students with an instructional technique that is considered time-consuming and arduous (e.g., Boulton & Leńko-Szymańska, 2015) and that has been described by students in higher education as tedious, laborious, time-consuming, and requiring extensive training (Chambers, 2007; see also Section 1.3). In the first article, it was found that the teachers' reluctance to have their students work with corpora was further exacerbated by their generally low opinions of their students' language proficiency and digital proficiency (Article 1). This worry is not unfounded, as teachers in a previous study observed that some of their students' low language proficiency was an obstacle to corpus consultation (Szudarski, 2020; see also Section 1.3). It can further be argued that the teachers' *perception of* the novelty space involved in DDL and the challenges it would pose for their learners were enough for the teachers to opt out of employing corpora at all. Hence, the novelty space was never addressed directly by the teachers but likely assumed to be too vast based on their own experiences with corpora.

Thus, the novelty problem has two distinct but connected layers. One layer is situated in practice and is perhaps the aspect to which the novelty space originally refers (cf. Sections 3.3.3 & 3.3.4). This layer is the actual newness that teachers and students face and the technological, pedagogical, and intellectual issues they must resolve when adapting to a new working method and a new tool. An example of such issues can be how to use a concordancer to find collocations or how to interpret corpus data. The previous studies cited above show that these issues have been identified before in pre-tertiary education through learner perspectives (see also Section 1.3), but as this section shows, the novelty problem is an amalgamation of issues related to tool design, new approaches, and new data, which should be addressed holistically. The second layer of the novelty problem is the teachers' *perception of novelty* and their knowledge and beliefs about how the new approach would succeed with their students. It is this second layer that may

keep corpora out of the classrooms altogether. It would seem that this obstacle related to the novelty problem is most efficiently handled during teacher training by designing university courses that introduce students to different types of corpora. However, a study with pre-service teachers by Leńko-Szymańska (2014) showed that a university course that gave an overview of corpus resources and applications over fourteen sessions was not sufficient to make them confident in introducing corpora into their practice. It therefore seems pertinent that these courses in addition should discuss corpora's pedagogic relevance and utility (see Sections 6.2.3 & 6.3.2) and give sound didactic guidance (see Section 6.3).

Some of these issues are not new. For example, the design of corpora and their content to better reflect learner needs (Braun, 2007) and ease of use has been discussed in the literature and solutions have been suggested such as pedagogic corpora (Hirata, 2020; Pérez-Paredes, 2020) or simpler interfaces such as SKELL [Sketch Engine for Language Learning]. However, the framing of these issues in the dissertation as the novelty problem contributes to our understanding of how these issues impact teachers and learners by suggesting that there are intersecting factors drawn across dimensions of technology, pedagogy, and content related to corpora (see Meunier, 2020), entrenched learning approaches, and curricular requirements that *exacerbate* the novelty problem in their amalgamation and might leave teachers and students with the feeling that corpora are competing for precious time as opposed to contributing to their writing and language learning. This issue, in turn, can influence the students' enthusiasm to brave the novelty space and invest in the new approach, as seemed to be the case in dissertation's case study (see Articles 2 & 3). In addition, teachers appear to be in part cognizant of this problem and are thus reluctant to increase the burden already put upon the students, which partially explains the research-practice gap. The main argument of this dissertation is that these issues can be handled in a more holistic manner through an inquiry framework, because this framework offers a focus on student-teacher relationships and differentiation. However, providing teachers with a way of guiding students through the novelty problem (see Section 6.3.1) is only one side of the coin. The other side entails giving them a reason to do so.

6.2.3 The Relevance Problem

The second core theme that emerged was how corpora and corpus-based approaches are relevant to Norwegian upper-secondary students and teachers. How corpus-based approaches can be integrated into the Norwegian upper-secondary classroom will rely on their relevance to

teachers, students, and the curriculum. Previous DDL studies and literature have shown aspects of DDL that may relate negatively to the issue of relevance. One such issue is the potential incongruence between the attention to fluency in communicative language teaching that permeates modern classrooms on the one hand, and the attention to accuracy in DDL on the other hand (Boulton & Leńko-Szymańska, 2015). Empirical support for this issue was found in the first article, as the teachers spoke of communicative competence as the central goal of the English subject. This goal is firmly in line with the communicative focus of both the old and new subject curricula (cf. Sections 2.3.2 and 2.3.4). In addition, the study found that the teachers tended to have a topic-centered focus in their lessons, and described language learning in terms of corrective feedback, immersion, and implicit acquisition through working with topics, while they mostly avoided grammar instruction (see Article 1). Thus, the relevance problem exists in part in the intersection between (a) teachers' pedagogical profile and epistemic ideas, which influence what they consider conducive to reaching curricular aims and how they structure their lessons, and (b) their perception of corpora's educational value. In other words, DDL might have too much of a language focus for a content-driven curriculum that prioritizes cultural and topical knowledge, learning strategies, and basic skills (cf. Section 2.3.2).

However, it was found that the teachers did do some explicit language instruction, especially through feedback (see Article 1). Of course, a focus on grammar and vocabulary is not an obstacle to the acquisition of communicative competence; rather, these are building blocks that underlie successful communication. The issue is not so much that the teachers did not believe in language instruction, but more one of how large a role it should play in their practice. As mentioned in the previous section, DDL is a demanding activity that introduces a lot of novelty requiring time and training. This investment may be seen as not worth it to a teacher who focuses on communication, culture, and the basic skills over an explicit focus on linguistic elements. Moreover, these observations suggest that DDL is not competing against traditional grammar teaching in the upper secondary classroom, but against communicative language teaching.

The case study attempted to bridge the gap between the teachers' topic-focused lessons and DDL by using the *Backbone* corpora with both topic- and language-centered annotation schemes, multimedia texts, and topic-appropriate language for younger learners, which made it a pedagogic corpus in line with the *feasibility scenario* (cf. Pérez-Paredes, 2020; Section 1.2.2). It should be remembered that the feasibility scenario is contrasted with the *possibilities scenario*, which denotes the use of established corpora and thus exploits the vast amount of

resources already available but intended for linguistics (ibid.). The feasibility scenario seeks relevance both in providing language from relevant topics and by supplying topic annotations in addition to metalinguistic ones. Theoretically, this shift toward more content-driven corpora would seem to fit the topic-focused lessons of the teachers well. However, a few obstacles were encountered when this tool was applied. Firstly, the students' feedback was that the topics were in fact not relevant, and they would have preferred topics closer to their personal interests (see Article 2). Secondly, the students did not see the teacher's introduction to the tool as sufficient and expected him to provide a clearer reason for working with it and to define its utility to language learning (see Article 3). The first issue can be connected to students' expectations of the content and aesthetics of digital applications, which are likely to be compared to commercial platforms that are continuously updated. The time required to create multimedia corpora that meet the requirements of the feasibility scenario, due to recording and transcribing texts and pedagogic tagging, leads to an *inertia* in the production of these types of resources. This inertia is an obstacle to relevance in that it delays the dissemination of the corpus from the time of data collection and thus makes it difficult to keep the tool current with respect to recent events, popular culture, or aesthetic features. The second issue shows that students expect to see the immediate utility of new applications to their needs. Similar findings were reported by Szudarski (2020), who found that some students had expectations of studying language directly relevant to their exam that were not met by the corpus approach. The need for instant applicability and intuitive tools is an obstacle to approaches which entail a high degree of novelty and whose learning benefits are delayed or not visible right away. For instance, there were several opportunities for metalinguistic and socio-cultural learning during the case study that were overlooked by the students who did not perceive achievement of learning outcomes arising from the corpus consultation (see Article 2). It also shows that the students' inclination to explore was not very pronounced, and that they expected the teacher to argue for the relevance of the approach.

Although an exception, a few students glimpsed some relevance in that *Backbone* sometimes provided them with patterns that showed how authentic English is structured. In addition, some of the students struggled with writing academic texts and wished for more concrete guidance from their teacher on text structure and vocabulary, such as sentence starters. The students said during the interviews that they had not and probably would not use *Backbone* again, but some of them had used *SKELL* [Sketch Engine for Language Learning], which had been introduced as an additional resource alongside *Backbone*, as a type of dictionary. Thus, the students had

found some utility in a more streamlined corpus resource and in the concordance lines from the *Backbone* that were not topic related. However, the students' concordance searches in the *Backbone* revealed another obstacle to the feasibility scenario that impacts its overall relevance. The extensive process of creating pedagogic corpora greatly limits the number of words in the corpus. Due to their aforementioned creation process, multimedia corpora are comprised of fewer and smaller texts, and consequently fewer words, than general corpora. For instance, students' searches led to very scant results in frequency lists and few concordance lines. The amount of data available in the *Backbone* made pattern interpretations through several concordance lines difficult, as some specific searches only yielded a few results, if any at all. Thus, corpus size is a major obstacle to the feasibility scenario because it diminishes one of the primary arguments for using corpora in education, namely analyzing language patterns across *multiple* examples from a rich data source. As the second article shows, the students were able to engage in a conversation around the concordance search of 'happy', but when searches of more specialized constructions were attempted, the lack of data in the corpus led to a swift stagnation in the activity (see Article 2).

However, as stated in the introduction, corpora can have many affordances, but the degree to which these affordances are realized depends on the ability and ingenuity of whoever is using them (Leńko-Szymańska & Boulton, 2015; cf. Section 1.2.1). The second article showed that sometimes these affordances come into play as a by-product of the task. One example was when students expressed their prejudices while engaging with Irish English (Article 2). Crucially, the teacher took the opportunity to address these prejudices through a series of pointed questions that prompted reflection and discussion. Such situations coupled with the teacher's presence of mind to tackle them show how pedagogic corpora can highlight types of speakers that are hidden in popular culture and how they can lead to instances of serendipitous learning, i.e., learning by happenstance (Bernardini, 2004), originally described as picking up linguistic constructions by luck, but equally applicable in these situations. Since a main utility of pedagogic corpora is to represent pedagogically neglected varieties of English (Pérez-Paredes, 2020), their affordance can be seen as supplying *representation of different speakers*, more so than *linguistic representativeness*. As was stated in the second article, one should "[expose] students to authentic language from non-American and non-British speakers to familiarize them with other varieties" (Karlsen, in preparation; Article 2). The provision of systematized and highlighted examples of pedagogically neglected varieties of English is a major utility of corpora that aim at representation of the many English varieties in the world and exemplifies

one way in which a corpus-based approach can be integrated in the classroom to promote the strengthened English-as-a-lingua-franca focus in the curriculum (cf. Section 2.3.4). It can also be argued that the multimedia nature of the *Backbone* enhanced this focus by hearing and seeing speakers of different varieties, which makes the differences more salient and further bolsters corpora as a tool for awareness raising of language variation (see Leńko-Szymańska & Boulton, 2015).

Lastly, it was concluded in the first article that the teachers might "[…] view language knowledge and thematic socio-cultural knowledge as different domains requiring different pedagogies" (Karlsen & Monsen, 2020, p. 142). The teachers tended to be more liberal toward student-active learning and inquiry-like processes while working with topics, while they appeared to have a stronger expert identity tied to language and gave more direct feedback on language form when the need arose. On the one hand, this means that there are already elements of inquiry-based education in their lessons such as examining different sources of information and making research questions. Similar elements were found in the case study (see Article 3). These elements can be building blocks for further inquiry-based approaches. On the other hand, it reveals potential obstacles to corpus integration due to the teachers' epistemic beliefs. Arguably, the teachers need to see the added value of DDL in addition to their general utility, i.e., how corpus-based approaches *enhance* learning. This added value is reinforced if seen in connection to the new curricular developments in Norway (see Section 2.3.3).

6.3 A Way Forward with Inquiry

The main research question has hitherto been addressed in terms of two central problems that emerged from the many obstacles that were identified in each of the three articles. These problems were broadly defined as *the novelty problem* and *the relevance problem* and were linked to the users' impressions and opinions. In the deliberation of these two core themes, it was suggested that these are two significant hurdles for successful DDL integration in upper-secondary school and a few solutions were sketched out. In the following, the primary theoretical contribution of this dissertation is discussed, namely the notion of DDL as a mode of inquiry (see Section 3.3) and how this new way forward addresses each of the aforementioned problems.

There is already considerable overlap in the learning mechanisms and activities of DDL and inquiry, but the latter has a clearer social dimension, which permeates its theoretical foundation and practical applications (cf. Section 3.3.2) and provides an incremental and diversified approach to inquiry pedagogy that involves teacher-student relationships and peer-to-peer learning. It is therefore argued that the novelty problem (Section 6.2.2) can be alleviated through this incremental approach and diversified educational role compositions, while DDL as inquiry finds new relevance in the new curriculum (Section 2.3.3). Although approaches to corpus consultation have been proposed that require minimal knowledge of corpora where the corpora are quick-and-easy reference works (e.g., Frankenberg-Garcia, 2014), what is proposed in this dissertation is a more comprehensive approach that satisfies the goal of the new core curriculum to promote in-depth learning (see Section 2.3.3). For instance, teaching students swiftly to check collocation when the need arises is a viable strategy and a simple way of getting corpora into the classroom; however, it is not conducive to in-depth learning and awareness-raising to the extent of an inquiry-based approach to DDL. DDL as *a mode of inquiry* builds on incrementality and diversification in order to alleviate the novelty problem (cf. Section 6.2.2); it acknowledges that inquiry is "a more-guided process based upon expertise from teachers and peers" (Saunders-Stewart et al., 2015), and "a *spectrum* of approaches, all of which are justified by common [socio-] constructivist principles" (Caputo, 2014, p. 371; Section 3.3.3). Consequently, DDL as a mode of inquiry requires an understanding of the teacher's expertise and the theoretical pragmatism (cf. Section 3.2.2) necessary to counteract the novelty problem and provides a framework that is highly relevant to the new curriculum in Norway. Each problem is addressed in turn in the following.

6.3.1 Addressing the Novelty Problem

In recent studies, increased attention has been given to the key role of the teacher or tutor in the success of direct applications of corpora with younger learners (Crosthwaite & Stell, 2020; Di Vito, 2020; Moon & Oh, 2018). Yet, little is still known about the broader picture of teachers' decision-making (Wicher, 2020), e.g. how teachers can facilitate and differentiate their approach to DDL or how teachers and students gradually adapt to and negotiate involvement and responsibility in a student-centered classroom. For instance, too much autonomy too soon can lead to students perceiving the teacher as passive or doubting his expertise (Karlsen, 2021; Article 3). This section furthers the recognition of the key role of the teacher and proposes

inquiry-based education as a way forward that holistically includes the teacher's expertise and emphasizes teacher-student relationships throughout the DDL process.

The subject of teacher expertise in DDL has been framed in terms of *corpus literacy*, which entails "a multicomponential set of complex skills" with four components:

1. Understanding basic concepts in corpus linguistics: What is a corpus and what types of corpora are available and how? What can you do – and cannot do [sic] – with a corpus?
2. Searching corpora and analysing corpus data by means of corpus software tools, e.g. concordancers: What is corpus software and how can it be used to search a corpus? How can corpus output be analysed?
3. Interpreting corpus data: How may general trends in language use/change be extrapolated from corpus data?
4. Using corpus output to generate teaching material and activities: How can you make use of corpus material for teaching purposes?

(Callies, 2019, p. 247)

The first three components are corpus-specific skills akin to those of a corpus linguist, while the fourth begins to touch on didactics. However, designing materials and activities are only one piece of the didactic puzzle that needs to be expanded upon if DDL is to support in-depth learning and corpora are to be more than reference works for students. In a classroom that strives for explorative, engaging, and creative approaches, it is necessary to train students in the three first components to work toward autonomy and deeper learning. Thus, teachers need the didactic knowledge and competence to guide students in the DDL process if they are to acquire new skills and become autonomous learners (e.g., Boulton & Cobb, 2017; Cheng, Warren, & Xun-feng, 2003; Johns, 1991; Millar & Lehtinen, 2008). It is argued that conceptualizing DDL as a mode of inquiry reinforces a social dimension missing from DDL (see Section 3.3.2) and one can begin to sketch out a *corpus didactics* as an expansion of the fourth component of corpus literacy. The notion of corpus didactics comes with its own components that are closely interlinked and co-dependent, and include educational roles, incrementality and differentiation, and scaffolding. These are discussed in the following.

Firstly, in order to facilitate DDL activities so that all learners can benefit from them without being overwhelmed, the teachers must guide students gradually into the new activities while keeping differentiation in mind. If providing "personalized help" to each student is essential in the approach's success (Moon & Oh, 2018; Section 1.3), issues will swiftly arise connected to situational factors such as time restraints, class size, and student level and competence disparity.

Studies such as Crosthwaite and Stell (2020), where the tutor was described as the key to the approach's success, only involved two students, but exemplified how they need different types of guidance (see also Section 1.3). However, one can look to inquiry-based education for a way of handling these issues pre-emptively that describes in more detail the deductive-inductive continuum sometimes referred to in DDL. Teachers can plan their lessons based on the types of inquiry proposed by Banchi and Bell (2008; Section 3.3.3) and prepare tasks that give various degrees of structure and information. For instance, one task might promote confirmation inquiry and have a formulated question, an outlined procedure, and a set solution, while another task is guided inquiry and only asks a question that students have to answer by choosing their own methods. In addition, prior to these tasks, the teacher can model the technique (Weimer, 2013), curate the data beforehand or pre-teach the target form (O'Keeffe, 2020) as a basis for confirmation inquiry. These are degrees of teacher mediation that can gradually move students toward more hypothetico-deductive and eventually inductive learning (cf. Section 3.3.3) and afford them an incremental increase in autonomy, self-regulation, and responsibility. They also align with proposed ways of scaffolding inquiry through gradually less support, teacher-provided guides, and templates (Hmelo-Silver et al., 2007; Section 3.3.3). However, the process does not need to be a unidirectional trajectory toward induction, as any of these tasks and techniques can be returned to or integrated simultaneously as a manner of level differentiation. These are examples in which early teacher-focused instructions and task preparation can begin to alleviate the novelty problem.

Secondly, beyond initial instruction and structured tasks, DDL activities can be scaffolded during lessons. Hmelo-Silver et al. (2007) suggest prompts for strategy use and teachers' providing expert information or guidance to the students among the scaffolding strategies (Section 3.3.3). In the case of DDL, these scaffolding strategies relate to the first three components of corpus literacy. This requires that teachers not only have tool-specific competency, such as the knowledge to demonstrate wildcard searches (e.g., Crosthwaite & Stell, 2020), but that they can model and guide students through processes of data analyses and interpretation as well. In order to provide these types of scaffolding, teachers need training in corpus linguistics that is coupled with their practical-pedagogic expertise so they can recognize and determine when and to whom each scaffolding strategy or task is appropriate. Some students only need the means to traverse the novelty space (i.e., an introduction to the tool), others need a map in the form of structured tasks and instructions, and others still might need to be guided every step of the way. Moreover, peer-to-peer assistance and teamwork have been

held up as positive experiences with DDL by some secondary school students (Di Vito, 2020) and peer scaffolding was observed in the case study where certain students had to explain the tool or process to other students (see Article 2). Thus, facilitating groupwork and collaborative learning opportunities can be a way in which teachers can encourage peer scaffolding (cf. Weimer, 2013; Section 3.3.4), and learners can take on the role of teacher or instructor and move toward greater responsibility in class while freeing up the teacher's time. However, one should also be aware of the potential pitfall of stronger students getting held up helping weaker students instead of progressing in their own learning toward more demanding tasks or forms of inquiry.

Thirdly, teachers must adapt to different roles in their practice and foster their students' journey toward a similarly pluralistic and flexible view of their own roles. The teacher in previous DDL literature has been described as a coordinator and director of student-initiated research (Johns, 1991), a facilitator (Gilquin & Granger, 2010), or a guide (Kennedy & Miceli, 2016), but with few detailed descriptions of what these roles entail. Moreover, as was argued in the third article of the dissertation, our understanding of roles in DDL should follow the concept of diversification from the inquiry literature so as to better understand the different roles one must take on in relation to different situations and problem (Karlsen, 2021; Article 3; see also Section 3.3.4). Central to the adoption of such a framework is a theoretical pragmatism that views teacher mediation as a key component in successful corpus integration and consequently refocuses on the position of the teacher in the DDL classroom. The roles students and teachers negotiate in the classroom will depend on their stage in role acquisition, their current proficiency, their current strategies, and the types of inquiry that are being explored.

For instance, one obstacle to corpus consultation is the lack of strategies that can be employed to analyze and interpret corpus output (e.g., Braun, 2007; Moon & Oh, 2018). On the other hand, the tutor's ability to model specific techniques spelled the success of concordance use with primary-school students in one study (Crosthwaite & Stell, 2020; see also Section 1.3). The teacher might adopt a more visible role in the initial stages so students can start exploring techniques and develop strategies for corpus searches and data analysis. In this *exploration* stage, the teacher could take on the role of the learner and model the whole process from hypothesis formulation through corpus search to data interpretation. In other words, the teacher solves the task while thinking out loud (Weimer, 2013; Section 3.3.4). This process should be prefaced by clear teacher instruction, which has been reported as a key teacher contribution in DDL (Di Vito, 2020). More student-centered tasks with varying degrees of scaffolding can then

be introduced, where the teacher must consider what types of inquiry different students should be pushed toward, who needs explicit instruction, and who simply needs a nudge in the right direction through prompts. During this process, the teachers should also examine their beliefs about students and whether they see them as capable of handling the novelty of the approach (Fiskum, Myhre, et al., 2018; Section 3.3.4). The teacher's decision to "[…] take risks and agree to 'let go' and let the student take pride of place in the classroom" (Gilquin & Granger, 2010, p. 367) will depend on their beliefs and the second layer of novelty outlined in Section 6.3.2.

Arguably, the novelty space can be tackled in the early stages of role *exploration* and *engagement*, which are the stages where considerable friction must be overcome before eventually *stabilization* and *diversification* of the new roles are reached (see Section 3.3.4). For instance, the smooth integration of DDL has been shown to be hindered by students' reliance on traditional, deductive methods (Szudarski, 2020) or just a reluctance to new embrace new types of instruction (Moon & Oh, 2018; see also Section 1.3). In some cases, inquiry requires that fossilized patterns of learning are "unlearned", which relies heavily on the student's adaptability, interests, attention span, and relationship to the teacher (Fiskum, Thorshaug, et al., 2018), and the same can be said of students engaging in DDL. However, students' reluctance to embrace DDL might be due to either reliance on more traditional, instructive grammar teaching, or the lack of explicit grammar teaching altogether. Moreover, in the dissertation's third article, it was reported that

> [t]he students expected the teacher to frame the lesson, i.e. *why* are they working with the corpus; specify the tool's usefulness, i.e. *what* can it be used for; motivate them; and instruct them on *how* it should be used once they got stuck, or even before the demand for aid arose. (Karlsen, 2021, p. 8)

These findings touch on other elements of the facilitator role, namely mutual understanding and shared responsibility (Doyle, 2011; Section 3.3.4). According to Fiskum, Thorshaug, et al. (2018), the inquiry teacher should strive for predictability in their expectations of their students (see Section 3.3.4). In fact, one of the students' criticisms during the case study was the unclear presence of the teacher and the occasional feeling that he "disappeared" from the process (see Article 3). Thus, it seems pertinent that the teacher both establishes a clear foundation for DDL by discussing and demonstrating its relevance to the students and remains *visible* throughout the process even when moving toward more autonomous working methods. As the discussion of scaffolding above shows, there are numerous ways in which the teacher can engage with

the student throughout their inquiry-based work, ranging from taking on the role of a learner by joining them to just prompting the appropriate strategy to use. Importantly, the teacher must not only rely on a strong initial stage but be prepared to revisit instruction or demonstrations with students throughout the process.

The above discussion sheds light on how the novelty problem associated with DDL can be tackled through a broadened understanding of teachers' decision making, student-teacher relationships, and the necessary view to a gradual and differentiated introduction of DDL to students, informed by inquiry-based education. These foci are best addressed during teacher education. If corpus courses are supposed to equip teachers with the necessary skills to instruct, facilitate, and guide student-active approaches to DDL, they require more than an overview of the available applications or an introduction to corpus analysis, they need the didactic training to do so. In addition, contributions can be made by the research field and DDL scholars. Since "[…] most instructors in DDL research are DDL scholars and not regular teachers" (Vyatkina, 2016, p. 207), little is still known about in-service, secondary school teachers' actions and reflections when integrating corpora into their everyday practice. What happens in cases where teachers, not scholars, are responsible for the preparation, planning, and implementation of DDL without direct researcher intervention is still relatively unknown. More qualitative studies would give us an idea of how the practical suggestions above would influence the DDL process and practitioners experience of said process.

6.3.2 Addressing the Relevance Problem

The relevance problem concerns the question of *why* Norwegian learners of English need DDL and *why* teachers should embrace it as an instructional technique or learning approach. A series of utilities of corpora to learners have been outlined in the literature that speak to their relevance to language learning. Corpora contain authentic language examples (Boulton & Cobb, 2017; Boulton & Leńko-Szymańska, 2015), which means that these examples are texts from English in use as opposed to examples constructed or engineered by textbook authors. Corpora also provide information on collocations, frequency, distribution, and context (Boulton, 2010), as well as language variety through genre and register metadata (Farr, 2008). The argument for pedagogical affordance is that working with authentic data leads to consciousness-raising (Boulton, 2020) or awareness-raising (Boulton & Leńko-Szymańska, 2015) about the varied

nature of languages, and that through DDL, access to corpus data affords an opportunity for an approach that promotes student autonomy and new learning skills (e.g., Boulton & Cobb, 2017; Cheng et al., 2003; Johns, 1991; Millar & Lehtinen, 2008). Moreover, meta-analyses of the effects of DDL on language acquisition have found evidence of positive outcomes (Boulton & Cobb, 2017; H. Lee, Warschauer, & Lee, 2017). Yet, there is little evidence of concordancers becoming normalized as an educational resource in the classrooms (Chambers, 2019), let alone in pre-tertiary classrooms in Norway (e.g., Kavanagh, 2021). Although these arguments and results speak to the utility of DDL as well as its potential to enhance language acquisition, the findings of the first article (Karlsen & Monsen, 2020; Article 1) alongside the general lack of normalization of corpus use in pre-tertiary contexts (Chambers, 2019) suggest that they are not sufficient to overcome the relevance problem. The relevance problem emerged in the discrepancy between the form and accuracy focus of DDL and the communicative and topic focus of the teachers, as well as the students' opinion of the tools relevance to their own learning process (see Section 6.2.3). The question of relevance can thus be divided into relevance for teachers and relevance for learners. These are addressed in the following, starting with the teacher perspective.

Meunier (2020) argues that DDL lacks constructive alignment, which entails the coherent and consistent alignment of the curriculum and curricular outcomes, the teaching methods, and the assessment tasks (p. 13). This issue resonates with the relevance problem in that the teaching and learning method of DDL poorly aligns – or is perceived by teachers to poorly align – with the communicative focus of the curriculum and the focus on (academic) writing and verbal skills on the exams. However, conceptualizing DDL as a mode of inquiry offers a holistic approach that aligns across the new curricular developments, the proposed teaching and learning methods, and assessment tasks. Firstly, the new subject curriculum has a renewed focus on linguistic forms and systems: "Language learning refers to developing language *awareness* and knowledge of English as a *system*, and the ability to use *language learning strategies*" (Utdanningsdirektoratet, 2020b; see also Section 2.3.4; my emphasis). This description incidentally aligns with the description of DDL's utility in its promotion of awareness-raising (Boulton, 2020; Boulton & Leńko-Szymańska, 2015), access to English patterns that show the variations of the system, and provides learners with new learning skills. Secondly, the social dimension of inquiry (cf. Section 3.3.2) can be emphasized in the classroom beyond teacher mediation by students engaging in collaborative learning (Caputo, 2014), communicating their results and newfound knowledge with their peers (Blessinger & Carfora, 2014), and facilitating

peer evaluation of student work (Weimer, 2013). For instance, students may come up with different answers to a research question due to the inherent variations in language, which they then in turn can discuss and make arguments for that are backed up by evidence. Tasks can also be designed that require topical and linguistic answers. For example, if students are working on Irish culture, one question in the task can be to explore Irish-English varieties. This blend of collaborative dialogue (cf. Section 3.2.2), inductive learning, and cultural topics seems primed to satisfy both communicative and linguistic goals while raising students' language awareness with regard to both linguistic and socio-cultural variation. Thirdly, the analytical and interpretive strategies required to work with corpus data (Callies, 2019), and the exploratory and student active involvement associated with inquiry (Caputo, 2014), fit with the call for scientific ways of working, critical thinking, problem-solving, investigative and analytical approaches, and assessment of information validity in the new curricular developments in Norway (Ludvigsen et al., 2015; see also Section 2.3.3). Fourthly, pedagogic corpora or corpora that represent different varieties of English including non-native representations, such as the *Backbone* corpora used in the case study (Articles 2 & 3), and the focus on speaker representation can be a window into English models that closer represent the English-as-a-lingua-franca [ELF] perspective that is part of the changing status of English in Norway (Rindal, 2020; see also Section 2.3.4). As reported in the third article, speaker representation via corpus data coupled with teacher scaffolding may lead to examination of prejudices through language form (Article 2), which is one way of reaching the curricular goal of promoting "[…] an exploratory approach to language, communicative patterns, lifestyles, ways of thinking and social conditions […]" (Utdanningsdirektoratet, 2020b, p. 2; see also Section 2.3.4).

In summary, DDL as a mode of inquiry has relevance to teachers in Norwegian upper-secondary schools because the approach fulfills curricular ambitions on four levels: (a) it provides a focus on language systems and the structure of English through (b) exploratory, engaging, and inquisitive approaches that (c) can be organized as academic presentations, collaborative learning, and knowledge sharing, which (d) opens up for types of Englishes that have typically been neglected in the educational system. It is therefore paramount that the curriculum is made a central document in corpus courses during teacher education, that the relevance of DDL to the curriculum is front and center, and that these courses' focus is not only on doing corpus linguistics but includes ways in which collaborative learning and knowledge sharing can be facilitated in a DDL setting. The issue can also be addressed by the research

community by conducting more research on peer-to-peer interactions during DDL activities as opposed to learner-to-corpus interactions.

The other side of the relevance coin is the students' opinions on the relevance of DDL to their own learning and interests. Recent studies in the pre-tertiary context show that students found DDL useful in helping them avoid grammar errors (Forti, 2020; Szudarski, 2020), finding authentic examples (Moon & Oh, 2018; Papaioannou et al., 2020), and increasing grammar consciousness and motivation for grammar learning (Moon & Oh, 2018). Positive opinions from the case study centered on a few students who thought seeing different varieties of English and how English was structured was somewhat interesting and relevant, but these positive aspects were overshadowed by a general lack of interesting topics and experiences of irrelevance (see Articles 2 & 3). Arguably, the time it took to learn the tool (i.e., the novelty space) and the lack of immediate utility (see Section 6.2.3) hindered the students' experience of the tool's usefulness. There are a few examples as to how this facet of the relevance problem could be alleviated. One example is to start with one function of the tool – i.e., concordance searches and analyses of concordance lines – and work in-depth with this function's utility before introducing others to make the path from introduction to utility shorter. Another example is simply that the teachers clarify the utility of the corpus to the students' needs through either modeling of the activity to demonstrate utility in relation to the problem at hand or discussing the utility with the students to engage them in the co-construction of knowledge; ideally, both these solutions are used concurrently. These solutions also highlight the teacher's role as a provider of context and direction, which are components the students in the case study felt were missing from the corpus implementation period (Article 3). They also open the door for critical evaluation of strategies and digital sources, which is a key object in both the core curriculum and the subject curriculum.

As discussed previously (Sections 1.2.3 & 6.2.3; see also Article 2), a solution to the issue of corpus relevance to pre-tertiary learners has been suggested through the feasibility scenario, that is the use of pedagogic corpora designed for topical relevance and topic- and level-appropriate language which are sensitive to students' needs (e.g., S. Braun, 2006; Hirata, 2020; Kohn et al., 2009; Pérez-Paredes, 2020). Whether teachers opt for the possibility scenario or the feasibility scenario depends on each scenario's utility and whether linguistic representation or representation of different speakers is the goal – note that pedagogic corpora can theoretically have many texts and provide linguistic representation of a particular type of speaker, but the ones that are available are relatively-speaking smaller in size than general corpora such as the

BNC. Whether corpora are introduced as quick-and-easy resources akin to reference works, used in DDL for inductive learning, or used to provide speaker representation is therefore contingent on what the teachers perceive their students' needs to be and whether they have the time, will and resources to make the necessary investments to learn the approach and the tools. In other words, the relevance of DDL grows out of how it can be utilized with different problems within the allotted timeframe and resources and is thus ultimately a pragmatic consideration.

Finally, it should be noted that what students *need* is not always what they *desire*, and the teacher in conjunction with the corpus data can be an *Other* who brings in something new (cf. Biesta, 2016; Section 3.2.2) in addition to providing structure and mediation (cf. O'Keeffe, 2020; Section 3.2.2). One might question whether the pursuit of topical relevance is worthwhile or a fool's errand. Creating corpora that have level- and language-appropriate texts definitely has its uses in that it provides some degree of data curation; however, chasing topical relevance in today's information society is an improbable ambition that will fall short due to the inevitable inertia in corpus design (see Section 6.2.3) and the accelerating perpetuity of knowledge production. The competition with the Web will lead to comparisons of relevance, utility, and aesthetics that can be unfortunate (see also Section 1.2.2, pp. 11-12, for a discussion of the web-as-corpus). It is therefore pertinent that the primary utilities and strengths of corpora are highlighted – both for teachers by teacher-educators, and for students by teachers – such as metadata, search replicability, cleaner data, and tagging and lemmatization (Boulton, 2015; see also Section 1.2.2), instead of attempting to compete with the video quality or updates of the Web.

What is more, the notion that corpus-based approaches should pander to students' interests and that these can fuel discovery learning (Bernardini, 2004; Hasselgård, 2014) assumes that students have linguistically directed interests. This assumption is faulty, or rather, the assumption that university students and pre-tertiary students have the same language interests is a false equivalence. Instead, one could perceive language interest as another goal and not a motivational driving force. The new core curriculum promotes students' curiosity and engagement (see Section 2.3.4), and the question then becomes how interest and curiosity might be cultivated in the learning environment. In the committee report leading up to the new Norwegian curriculum, it was stated that "young people are by nature inquisitive and exploring, but curiosity must be stimulated to be developed" (Ludvigsen et al., 2015, p. 33; see also Section 2.3.3). In other words, subject-specific curiosity should not be assumed a priori, but be carefully

cultivated. According to Andersen et al. (2018), the focus on exploration, creativity and engagement in the new curriculum should in part be pursued by cultivating *a sense of wonder* that fosters students' curiosity and helps students pose more questions rather than answering them (see also Section 3.3.1). Another way interests and curiosities may be cultivated is through the social dimension of inquiry, i.e., collaborative learning, sharing new knowledge with their peers, and project-based learning (Blessinger & Carfora, 2014; Caputo, 2014; see also Section 3.3.2). This is not to say that one should not be cognizant of students' interests and include them in the learning process, but equally one should avoid the notion of student-as-consumer whose needs must be met at every turn (Biesta, 2016; see also Section 3.2.2). Instead, DDL as inquiry opens up for a renewed focus on teacher-student relationships through teacher mediation, instruction and guidance, and collaborative processes and knowledge sharing in the classroom that bring something new and hopefully serve to cultivate interest and curiosity in the students.

6.4 Contributions, Limitations, and Future Directions

The primary empirical contribution from the first research phase was the data on teacher beliefs and how they intersected with students' experiences and opinions. These data re-affirmed obstacles to DDL from previous research but in a Norwegian setting, as well as adding new perspectives from teachers who, crucially, were both corpus-trained and in-service. The data also shed light on how converging factors related to technology, corpus content, and pedagogy contribute to the barriers teachers experience when considering a corpus-based approach. Theoretical contributions were made by introducing the TPACK framework (Koehler & Mishra, 2005), and teachers' beliefs and technology integration (Kim, Kim, Lee, Spector, & DeMeester, 2013) as descriptive categories for data analyses. The factors have been highlighted in other DDL studies (Leńko-Szymańska, 2017; Meunier, 2020), but the focus on teachers' beliefs as a barrier to pedagogic corpus integration introduced a new category against which to identify the challenges DDL faces in pre-tertiary education; it therefore signals a new theoretical contribution to the field. The methodological contribution of the first phase was the semi-structured interviews and questionnaire that both followed the same design principle with categories from the aforementioned theoretical perspectives and moved from open-ended questions about pedagogic and digital experience and opinions toward more pointed questions about corpus experience and opinions. The design of these instruments allowed for comparisons

between student and teacher answers and the uncovering of ancillary or influential factors that impacted corpus consultation such as digital competence or learner preferences.

The second research phase contributed empirical data on students' experience with and opinion of a pedagogic corpus tool and the process of learning to use it, as well as on teacher-student interactions and students' expectations. These data revealed further obstacles to Norwegian students' corpus consultation but also opportunities of learning in an authentic pre-tertiary context. The theoretical contribution of this research phase was the application of the feasibility scenario and pedagogic corpora (Pérez-Paredes, 2020) and the application of socio-cultural concepts such as collaborative dialogue and peer and teacher mediation (cf. O'Keeffe, 2020; Section 3.2.2). The methodological contributions include lesson plan and task design, where the tasks were structured guides on how to use the *Backbone website* and involved teacher-researcher collaborations. This approach chosen was in response to the criticism that few in-service teachers conduct DDL research (Vyatkina, 2016).

The contributions from both research phases culminated in the present chapter's two categories, *the novelty problem* and *the relevance problems*, as theoretical and potential analytical categories for future DDL research that encompass the many challenges related to training, beliefs, perceptions of relevance, and relevance to the local level (i.e., relevance to the curriculum). Meanwhile, the addition of an inquiry-based framework is a theoretical contribution that proposes a more systematic manner of approaching planning and scaffolding in DDL and ways in which communicative and exploratory perspectives can both find a place in the classroom through collaborative learning, co-construction and sharing of knowledge, and peer scaffolding.

There are several limitations to the dissertation's studies and conclusions. The first research phase involved only four teachers and their students in a total of four schools, while the second phase only looked at one teacher and his students. The small sample size, particularly in the case study, means that the opinions and perspectives of the participants are prone to idiosyncrasies and situational/contextual factors. The results should be considered in relation to other findings of a similar character from other contexts and future research could apply the same methodological and theoretical concepts with upper secondary school students from different schools and areas.

Another limitation is in the time and focus of the case study (see Chapter 4). Only two weeks were scheduled for the implementation of corpora, and although structured guides and an

allegedly intuitive tool were utilized, the novelty of the approach may have required more training and only allowed the students and teacher to scratch the surface of what was possible to achieve. In addition, the teacher expected the researcher to provide lesson plans and tasks, which means that he was less involved in the planning process than intended. Time constraints are the reality of a lot of classrooms and in-service teachers are often protective of their time, which makes getting access a challenge. Nevertheless, a research project where (1) the teacher-research collaboration is more balanced or the teacher does all the planning, and (2) more time is allotted to the implementation period so that an incremental and differentiated approach such as the one described in this introductory chapter can be implemented would likely yield data that represent the classroom experience more authentically and that show students overcoming the first hurdle of training and familiarization with the new tools. Alternatively, a narrower approach that focuses on a specific function of a particular corpus could be implemented.

A final limitation is the theoretical nature of the inquiry-based approach to DDL. Although the case study entailed several social components such as groupwork and collaborative dialogue, the more systematic idea of DDL as a mode of inquiry grew out of the negatively skewed opinions and experiences of the students to the corpus-integration period. According to Pérez-Paredes (2020), there are few studies on the topic of corpus consultation in pre-tertiary education. An interesting way forward would be longitudinal studies that report on the systematic, gradual introduction of DDL as a mode of inquiry and that focus on student-teacher and peer-to-peer processes alongside learner-to-corpus interactions and outcomes.

6.5 Concluding Remarks

This dissertation has investigated the use of corpus-based approaches in pre-tertiary education through teacher and student perspectives and a case study. The main research question it sought to answer was: *How can corpus-based approaches be integrated into Norwegian secondary schools and how are they received by the users?* As the discussion shows, there were several obstacles which hindered this integration and that must be overcome for direct applications of corpora to become normalized in the classroom (see Chambers, 2019 for a discussion of the nomarlization of corpus consultation in schools). These obstacles emerged from both previous studies and the dissertation's contributions, which broadly fell into the categories of *novelty* and *relevance*. It was further argued that these obstacles can be addressed by conceptualizing and practicing DDL as a mode of inquiry. Inquiry-based education adds a social component to an

otherwise individualistic, constructivist DDL that places emphasis on the teacher's role in DDL and refocuses on the social processes of learning through collaborative learning and peer-to-peer interactions.

The dissertation has shown that there are still some major obstacles to the successful and smooth integration of DDL in secondary schools, but also that DDL both in utility and working procedures coincides extraordinarily well with the exploratory, inquisitive, critical, and student active ambitions in the new Norwegian curriculum (cf. Section 2.3.3). Thus, DDL can find a new home in the Norwegian curricular and subject renewal if it is framed in terms of its value as an inquiry-based approach. Simultaneously, the new focus on in-depth learning and research-emulating working methods in the new curriculum gives room for and requires more demanding instructional techniques and learning processes, which means that teachers are now justified more than ever in dedicating time and resource in order to facilitate corpus-based approaches.

As stated in Section 2.3.3, DDL as a mode of inquiry becomes one way in which (1) subject-specific, scientific ways of working can enter the language classroom, and (2) the cross-curricular ambitions of the new curriculum to promote transferable, research-like skills and different strategies can be worked at in the English classroom. Corpus linguistics is after all a major branch of empirical science within linguistics. It also highlights the fact that language, our foremost cultural artefact, and the medium through which we mediate our experiences, is in itself varied, complex, and socially situated, and ripe for investigation in exciting new ways. Furthermore, the skills and principles of inquiry that coincide with the core curriculum should be working their way into Norwegian schools on every level in the years to come. What this means is that many of the issues with the novelty problem, such as inductive reasoning or hypothesis-making, are tackled in more than one subject, which may reduce the novelty space altogether and help pave the way for DDL as a mode of inquiry in the subject English classroom. Thus, Norwegian school researchers, teacher-educators, and teachers can find common ground in the curricular goals of exploratory, inquisitive, and engaging goals across disciplines and subjects. The use of corpora in secondary school remains "[…] relatively uncharted territory" (Wicher, 2020), but this dissertation has voyaged into this territory and begun to chart a way forward.

References

Andersen, H. P., Fiskum, T. A., & Rosenlund, M. R. (2018). Hva menes med undrende, utforskende og aktiviserende undervisning? In T. A. Fiskum, D. Gulaker, & H. P. Andersen (Eds.), *Den engasjerte eleven : undrende, utforskende og aktiviserende undervisning i skolen* (pp. 17-30). Oslo: Cappelen Damm akademisk.

Anthony, L. (2015). TagAnt (Version 1.1.0) Tokyo, Japan: Waseda University. Retrieved from https://www.laurenceanthony.net/software

Anthony, L. (2019). AntConc (Version 3.5.8). Tokyo, Japan: Waseda University. Retrieved from www.laurenceanthony.net/software

Aulls, M. W., Tabatabai, D., & Shore, B. M. (2016). What Makes Inquiry Stick? The Quality of Preservice Teachers' Understanding of Inquiry. *SAGE Open, 6*(4), 215824401668139. doi:10.1177/2158244016681394

Banchi, H., & Bell, R. (2008). THE MANY LEVELS OF Inquiry. *Science and children, 46*(2), 26-29.

Bereiter, C. (1994). Constructivism, Socioculturalism, and Popper's World 3. *Educational researcher, 23*(7), 21-23. doi:10.3102/0013189X023007021

Bergsli, H. (2013). Helse og frafall i videregående opplæring: Underlagsrapport til Sosial ulikhet i helse: En norsk kunnskapsoversikt. Retrieved from http://www.forebygging.no/Global/Bergsli%20Helse%20og%20frafall%20i%20videreg%C3%A5ende%20oppl%C3%A6ring%20(2).pdf

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future development. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Philadelphia: J. Benjamins.

Biesta, G. (2016). *The beautiful risk of education*. In Interventions: education, philosophy, and culture.

Bjørndal, C. R. P. (2013). *Det vurderende øyet: Observasjon, vurdering og utvikling i undervisning og veiledning* (2nd ed.). Oslo: Gyldendal Akademisk.

Blessinger, P., & Carfora, J. M. (2014). Innovative approaches in teaching and learning: An introduction to inquiry-based learning for the arts, humanities, and social sciences In J. M. Carfora & P. Blessinger (Eds.), *Inquiry-Based Learning for the Arts, Humanities and Social Sciences: A Conceptual and Practical Resource for Educators* (Vol. 2, pp. 3-25). Bingley: Bingley: Emerald Publishing Limited.

Bloomfield, L. (1935). *Language*. London: Allen & Unwin.

Boeije, H. (2010). *Analysis in qualitative research*. Los Angeles: SAGE.

Boulton, A. (2009). Testing the limits of data-driven learning: language proficiency and training. *ReCALL (Cambridge, England), 21*(1), 37-54. doi:10.1017/S0958344009000068

Boulton, A. (2010). Data-Driven Learning: Taking the Computer Out of the Equation. *Language Learning, 60*(3), 534-572.

Boulton, A. (2015). Applying data-driven learning to the web. In A. Lénko-Szymánska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 267-295). Amsterdam/Philadelphia: John Benjamins publishing company.

Boulton, A. (2020). Foreword. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. London: Routledge* (pp. xiii-xx). Oxon: Abingdon: Routledge.

Boulton, A., & Cobb, T. (2017). Corpus Use in Language Learning: A Meta-Analysis. *Language Learning, 67*(2), 348-393.

Boulton, A., & Leńko-Szymańska, A. (2015). Introduction: Data-driven learning and language pedagogy. In A. Leńko-Szymańska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 1-14). Amsterdam/Philadelphia: John Benjamins publishing company.

Braun, S. (2006). ELISA–a pedagogically enriched corpus for language learning purposes. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus Technology And Language Pedagogy: New Resources, New Tools, New Methods* (pp. 25-47). Frankfurt/M: Lang.

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

Brenner, M. E. (2006). Interviewing in Educational Research. In *Handbook of complementary methods in education research* (pp. 357-370). Mahwah, N.J. Washington, D.C.: Lawrence Erlbaum Associates; Published for the American Educational Research Association.

Brevik, L. M. (2019). Gamers, Surfers, Social Media Users: Unpacking the role of interest in English. *Journal of Computer Assisted Learning, 35*(5), 595-606. doi:10.1111/jcal.12362

Brevik, L. M., Skarpaas, K. G., & Isaksen, A. R. (2020). Vocational English: Building covational orientation and relevance. In L. M. Brevik & U. Rindal (Eds.), *Teaching English in Norwegian Classrooms: From research to practice*. Oslo: Universitetsforlaget.

Brinkmann, S., & Kvale, S. (2015). *InterViews: Learning the Craft of Qualitative Research Interviewing* (3rd ed.). Los Angeles: Sage.

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. vi, 267 pages). Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Caputo, L. (2014). Using inquiry-based learning to teach additional language in a high school context. In J. M. Carfora & P. Blessinger (Eds.), *Inquiry-Based Learning for the Arts, Humanities and Social Sciences: A Conceptual and Practical Resource for Educators* (Vol. 2, pp. 369-391). Bingley: Bingley: Emerald Publishing Limited.

Chambers, A. (2007). Popularising corpus consultation by language learners and teachers. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the Foreign Language Classroom* (Vol. 61, pp. 1-16). Amsterdam: BRILL.

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Lang. Teach, 52*(4), 460-475. doi:10.1017/S0261444819000089

Charles, M. (2014). Getting the corpus habit: EAP students' long-term use of personal corpora. *English for Specific Purposes, 35*, 30-40. doi:10.1016/j.esp.2013.11.004

Checkbox.com. (2018). Checkbox (Version 2018Q2). Massachusetts, US: Checkbox Survey Inc. Retrieved from https://www.checkbox.com/

Cheng, W., Warren, M., & Xun-feng, X. (2003). The language learner as language researcher: putting corpus linguistics on the timetable. *System, 31*(2), 173-186. doi:http://dx.doi.org/10.1016/S0346-251X(03)00019-8

Chomsky, N. (1953). Systems of Syntactic Analysis. *The Journal of Symbolic Logic, 18*(3), 242-256.

CLARIN. (2012). CLARINE ERIC. Retrieved from https://www.clarin.eu/portal

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (Eighth edition. ed.). London ; New York: Routledge.

The British National Corpus. (2007). The British National Corpus, version 3 (BNC XML Edition). BNC XML Edition. Retrieved from http://www.natcorp.ox.ac.uk/

National Research Council. (1996). *National science education standards: Observe, interact, change, learn*. Washington, DC: National Academies Press.

Crosthwaite, P. (2020a). Data-Driven Learning and Younger Learners: Introduction to the volume. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. London: Routledge* (pp. 1-10). Oxon: Abingdon: Routledge.

Crosthwaite, P. (2020b). *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. London: Routledge*.

Crosthwaite, P., & Stell, A. (2020). "It helps me get ideas on how to use my words": primary school students' initial reactions to corpus use in a private tutoring setting. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners.* London: Routledge.

Crystal, D. (2012). *English as a Global Language*.

Dalton-Puffer, C. (2014). *Corpus Linguistics in Language Teacher Education. Plenary lecture given at the 14th Klagenfurt Conference on Corpus Based Applied Linguistics (CALK 14)*. Paper presented at the Klagenfurt Conference on Corpus Based Applied Linguistics (CALK 14), University of Klagenfurt, Austria.

Davies, M. (2008). The Corpus of Contemporary American English. Retrieved from https://www.english-corpora.org/coca/

De Bartolo, A. M. (2020). A Preliminary Exploration of Teachers' Attitudes towards Intercultural Communication through ELF. *EL.LE, 8*(3). doi:10.30687/ELLE/2280-6792/2019/03/006

Di Vito, S. (2020). Teaching French to young learners through DDL. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. London: Routledge.

Doyle, T. (2011). *Learner-centered teaching: putting the research on learning into practice* (1st ed ed.). Sterling, Va: Sterling, Va: Stylus Pub.

Ellis, N. C. (2019). Essentials of a Theory of Language Cognition. *Modern Language Journal, 103*(S1), pl. doi:10.1111/modl.12532

Ellis, N. C., & Wulff, S. (2015). Usage-Based Approaches to SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition : an introduction* (Second Edition. ed., pp. 75-93). New York: Routledge.

Engelsen, B. U., & Karseth, B. (2007). Læreplan for Kunnskapsløftet - et endret kunnskapssyn? *Norsk pedagogisk tidskrift, 91*(5), 404-415.

Evans, V., & Green, M. (2006). *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh: Edinburgh University Press.

Farr, F. (2008). Evaluating the Use of Corpus-based Instruction in a Language Teacher Education Context: Perspectives from the Users. *Language Awareness, 17*(1), 25-43. doi:10.2167/la414.0

First, E. (2020). EF English Proficiency Index. Retrieved from https://www.ef-danmark.dk/assetscdn/WIBIwq6RdJvcD9bc8RMd/legacy/__/~/media/centralefcom/epi/downloads/full-reports/v10/ef-epi-2020-english.pdf

Fiskum, T. A., Myhre, H., & Rosenlund, M. R. (2018). Lærerens valg og holdninger: Hvor er jeg? In T. A. Fiskum, D. Gulaker, & H. P. Andersen (Eds.), *Den engasjerte eleven : undrende, utforskende og aktiviserende undervisning i skolen* (pp. 31-44). Oslo: Cappelen Damm akademisk.

Fiskum, T. A., Thorshaug, A., & Husby, J. A. (2018). Eleven i fokus. In T. A. Fiskum, D. Gulaker, & H. P. Andersen (Eds.), *Den engasjerte eleven : undrende, utforskende og aktiviserende undervisning i skolen* (pp. 45-67). Oslo: Cappelen Damm akademisk.

Fladmoe, A. (2013). Education in the News and in the Mind. *Nordicom review, 33*(1), 99-116. doi:10.2478/nor-2013-0007

Flowerdew, L. (2015). Data-driven learning and language learning theories: Whither the twain shall meet. In A. Lénko-Szymánska & A. Boulton (Eds.), *Multiple affordances of language corpora for data-driven learning* (pp. 15-36). Amsterdam/Philadelphia: John Benjamins publishing company.

Forti, L. (2020). Evaluating Data-Driven Learning Effects in the Italian L2 Classroom: Etic and Emic Perspectives Combined. *EL.LE*(2). doi:10.30687/ELLE/2280-6792/2019/02/006

Franc, C., & Morton, A. (2014). Inquiry-based learning for language learning: The case of French advanced level at the university of Manchester (UK). In J. M. Carfora & P. Blessinger (Eds.), *Inquiry-Based Learning for the Arts, Humanities and Social Sciences: A Conceptual and*

Practical Resource for Educators (Vol. 2, pp. 77-103). Bingley: Bingley: Emerald Publishing Limited.

Frankenberg-Garcia, A. (2014). The use of corpus examples for language comprehension and production. *ReCALL, 26*(2), 128-146. doi:10.1017/S0958344014000093

Frankenberg-Garcia, A. (2014). How language learners can benefit from corpora, or not. *Recherches en didactique des langues et des cultures [En ligne], 11*(1), 1-8.

Friginal, E. (2018). *Corpus linguistics for English teachers : new tools, online resources, and classroom activities*. New York, NY: Routledge.

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (1st ed., pp. 359-370). London ; New York, NY: Routledge.

Harmer, J. (2015). *The practice of English language teaching* (5th ed. ed.). Harlow: Pearson Education Longman.

Hasselgård, H. (2014). Å bruke korpus i språkundervisningen. In C. Bjørke, M. Dypedahl, & G. A. Myklevold (Eds.), *Fremmedspråksdidaktikk* (pp. 139 - 156): Cappelen Damm Akademisk.

Haugsbakk, G. (2013). From Sputnik to PISA Shock - New Technology and Educational Reform in Norway and Sweden. *Education enquiry, 4*(4), 23222. doi:10.3402/edui.v4i4.23222

Hirata, E. (2020). The Development of a Multimodal Corpus Tool for Young EFL Learners. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. London: Routledge* (pp. 88-105). Oxon: Abingdon: Routledge.

Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99-107. doi:10.1080/00461520701263368

Hølleland, H. (2007). Innføring i Kunnskapsløftet. In H. Hølleland (Ed.), *På vei mot Kunnskapsløftet: Begrunnelser, løsninger og utfordringer*. Oslo: Cappelen Akademisk Forlag.

ICE-GB. (2020). A brief history of the Survey of English Usage. Retrieved from https://www.ucl.ac.uk/english-usage/about/history.htm

Isnes, A. (2007). Læreplanverket i Kunnskapsløftet. In H. Hølleland (Ed.), *På vei mot Kunnskapsløftet: Begrunnelser, løsninger og utfordringer* (pp. 84-111). Oslo: Cappelen Akademisk Forlag.

Johns, T. (1991). "Should you be persuaded": Two samples of data-driven learning materials. *Classroom Concordancing ELR Journal, 4*, 1-16.

Jonassen, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational technology research and development, 39*(3), 5-14. doi:10.1007/bf02296434

Karlsen, P. H. (2021). Educational Roles in Corpus-Based Education: From Shift to Diversification. *Nordic Journal of Language Teaching and Learning, 9*(1).

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.

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.

Keiser, M. M., Burrows, B. D., & Randall, B. (2014). Growing teachers for a new age: Nurturing constructivist, inquiry-based teaching. In J. M. Carfora & P. Blessinger (Eds.), *Inquiry-Based Learning for the Arts, Humanities and Social Sciences: A Conceptual and Practical Resource for Educators: Innovations in higher education teaching and learning* (Vol. 2, pp. 303-324). Bingley: Bingley: Emerald Publishing Limited.

Kennedy, C., & Miceli, T. (2016). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning, 30*(1-2), 91-114. doi:10.1080/09588221.2016.1264427

Kilgarriff, A., & Grefenstette, G. (2003). Introduction to the Special Issue on the Web as Corpus. *Computational linguistics - Association for Computational Linguistics, 29*(3), 333-347. doi:10.1162/089120103322711569

Kim, C., Kim, M. K., Lee, C., Spector, J. M., & DeMeester, K. (2013). Teacher beliefs and technology integration. *Teaching and Teacher Education, 29*, 76-85. doi:10.1016/j.tate.2012.08.005

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist, 41*(2), 75-86. doi:10.1207/s15326985ep4102_1

Koehler, M. J., & Mishra, P. (2005). WHAT HAPPENS WHEN TEACHERS DESIGN EDUCATIONAL TECHNOLOGY? THE DEVELOPMENT OF TECHNOLOGICAL PEDAGOGICAL CONTENT KNOWLEDGE. *Journal of Educational Computing Research, 32*(2), 131-152.

Kohn, K., Hoffstaedter, P., & Widmann, J. (2009). *BACKBONE - Pedagogic Corpora for Content & Language Integrated Learning*. Paper presented at the EUROCALL 2009 Conference, Universidad Politecnica de Valencia.

Lantolf, J. P., Thorne, S. L., & Poehner, M. E. (2015). Sociocultural Theory and SLA. In B. VanPatten & J. Williams (Eds.), *Second Language Acquisition Research Series* (Second Edition. ed., pp. 207-226). New York: Routledge.

Lee, H., Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning & Technology, 21*(2), 32-51.

Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the Establishment of a Data-Driven Learning Model: Role of Learner Factors in Corpus-Based Second Language Vocabulary Learning. *The Modern language journal (Boulder, Colo.), 104*(2), 345-362. doi:10.1111/modl.12634

Lee, J. H., Lee, H., & Sert, C. (2015). A Corpus Approach for Autonomous Teachers and Learners: Implementing an On-Line Concordancer on Teachers' Laptops. *Language Learning & Technology, 19*(2), 1.

Leech, G. (1997). Teaching and language corpora: A convergence. . In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 11-23). Harlow: Addison Wesley Longman.

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. *26*(2), 260-278. doi:10.1017/S095834401400010X

Leńko-Szymańska, A. (2017). Training Teachers in Data-Driven Learning: Tackling the Challenge. *Language Learning & Technology, 21*(3), 217-214.

Leńko-Szymańska, A., & Boulton, A. (2015). Introduction: Data-driven learning in language pedagogy. In A. Lenko-Szymanska & A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-Driven Learning* (pp. 131-154). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lillejord, S., Halvorsrud, K., Ruud, E., Morgan, K., Freyr, T., Fischer-Griffiths, P., . . . Manger, T. (2015). Frafall i videregående opplæring: En systematisk kunnskapsoversikt. Retrieved from https://utdanningsforskning.no/globalassets/frafallrapport2015-kunnskapssenteret.pdf

Linguistics, C. f. E. C. (2019). Learner Corpora around the World. Retrieved from https://uclouvain.be/en/research-institutes/ilc/cecl/learner-corpora-around-the-world.html

Liontou, T. (2020). The effect of data-driven learning activities on young EFL learners' processing of English idioms. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. London: Routledge.

Liu, D., & Lei, L. (2017). *Using Corpora for language learning and teaching*. Alexandria, VA: TESOL International Association.

Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014: designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics, 22*(3), 319-344.

Ludvigsen, S., Gundersen, E., Indregard, S., Ishaq, B., Kleven, K., Korpås, T., . . . Øye, H. (2015). The School of the Future: Renewal of subjects and competences. Retrieved from https://www.regjeringen.no/contentassets/da148fec8c4a4ab88daa8b677a700292/en-gb/pdfs/nou201520150008000engpdfs.pdf

Markussen, E. (2007). Reform 94 lever videre - men videregående svikter hver femte ungdom! . In H. Hølleland (Ed.), *På vei mot Kunnskapsløftet: Begrunnelser, løsninger og utfordringer* (pp. 84-111). Oslo: Cappelen Akademisk Forlag.

McEnery, T., & Xiao, R. (2010). *Corpus-Based Contrastive Studies of English and Chinese* (1 ed. Vol. 11). London: London: Routledge.

McEnery, T., & Xiao, R. (2011). What Corpora Can Offer in Language Teaching and Learning. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years On* (pp. 364-380). Amsterdam/New York: Rodopi.

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

Meunier, F. (2020). A case for constructive alignment in DDL. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. London: Routledge.

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

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. *Journal of Research in Science Teaching, 47*(4), 474-496. doi:10.1002/tea.20347

Moon, S., & Oh, S.-Y. (2018). Unlearning overgenerated be through data-driven learning in the secondary EFL classroom. *ReCALL, 30*(1), 48-67. doi:10.1017/S0958344017000246

Mukherjee, J. (2006). corpus linguistics and language pedagogy: The state of the art - and beyond. In S. Braun, K. Kohn, & J. Mukherjee (Eds.), *Corpus technology and language pedagogy : new resources, new tools, new methods* (pp. 5-24). New York: P. Lang.

Munden, J., & Sandhaug, C. (2017). *Engelsk for secondary school*. Oslo: Gyldendal akademisk.

O'Keeffe, A. (2020). Data-driven learning – a call for a broader research gaze. *Language Teaching*, 1-14. doi:10.1017/S0261444820000245

O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL, 19*(3), 269-286. doi:10.1017/S095834400700033X

Palinkas, L., Horwitz, S., Green, C., Wisdom, J., Duan, N., & Hoagwood, K. (2015). Purposeful Sampling for Qualitative Data Collection and Analysis in Mixed Method Implementation Research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533-544. doi:10.1007/s10488-013-0528-y

Papaioannou, V., Mattheoudakis, M., & Agathopoulou, E. (2020). The effect of data-driven learning in a Greek secondary education setting: the implementation of a blended approach. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. London: Routledge.

Pérez-Paredes, P. (2020). The Pedagogical 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. London: Routledge*. Oxon: Abingdon: Routledge.

Rindal, U. (2014). What is English? *Acta Didactica Norge, 8*(2), 1-17.

Rindal, U. (2020). English in Norway: A language and a subject in transition. In L. M. Brevik & U. Rindal (Eds.), *Teaching English in Norwegian classrooms: From research to practice* (pp. 23-42). Oslo: Universitetsforlaget.

Römer, U. (2011). Corpus Research Applications in Second Language Teaching. *Annual Review of Applied Linguistics, 31*, 205-225. doi:10.1017/S0267190511000055

Saunders-Stewart, K. S., Gyles, P. D. T., Shore, B. M., & Bracewell, R. J. (2015). Student outcomes in inquiry: students' perspectives. *Learning environments research, 18*(2), 289-311. doi:10.1007/s10984-015-9185-2

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction*. Cambridge: Cambridge University Press.

Schreier, M. (2018). Sampling and Generalization. In U. Flick (Ed.), *The Sage handbook of qualitative data collection* (pp. 84-98). Los Angeles: Sage Reference.

Simpson, R. C., Briggs, S. L., Ovens, J., & Swales, J. M. (2002). The Michigan Corpus of Academic Spoken English.

Sinclair, J. M. (1997). Corpus Evidence in Language Description. In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 27-39). Harlow: Addison Wesley Longman.

Smith, M. L. (2006). Multiple Methodology in Education Research. In J. L. Green, G. Camilli, P. B. Elmore, & American Educational Research Association. (Eds.), *Handbook of complementary methods in education research* (pp. 457-475). Mahwah, N.J. Washington, D.C.: Lawrence Erlbaum Associates; Published for the American Educational Research Association.

Strike, K. A. (2006). The Ethics of Educational Research. In J. L. Green, G. Camilli, P. B. Elmore, & A. E. R. Association. (Eds.), *Handbook of complementary methods in education research* (pp. xxix, 865 p.). Mahwah, N.J. Washington, D.C.: Lawrence Erlbaum Associates; Published for the American Educational Research Association.

Szudarski, P. (2020). Effects of data-driven learning on enchancing the phraseological knowledge of secondary school learners of L2 English. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners.* London: Routledge.

Thomas, S. L. (2006). Sampling: Rationale and Rigor in Choosing What to Observe. In C. F. Conrad & R. C. Serlin (Eds.), *The SAGE Handbook for Research in Education* (pp. 393-404). Thousand Oaks, California: SAGE Publications, Inc.

Utdanningsdirektoratet. (2019). Søkertall videregående opplæring for skoleåret 2019-2020 (analyse). Retrieved from https://www.udir.no/tall-og-forskning/statistikk/statistikk-videregaende-skole/sokertall-videregaende-opplaring-for-skolearet-2019-2020/

Utdanningsdirektoratet. (2013). English subject curriculum (ENG1-03) Retrieved from https://www.udir.no/kl06/ENG1-03?lplang=http://data.udir.no/kl06/eng

Utdanningsdirektoratet. (2020a). Core curriculum – values and principles for primary and secondary education. Retrieved from https://www.udir.no/lk20/overordnet-del/?lang=eng

Utdanningsdirektoratet. (2020b). Curriculum in English (ENG01-04). Retrieved from https://www.udir.no/lk20/eng01-04?lang=eng

Utdanningsdirektoratet. (2020c). Hva er fagfornyelsen? Retrieved from https://www.udir.no/kl06/ENG1-03?lplang=http://data.udir.no/kl06/eng

Utdanningsdirektoratet. (2020d). Utdanningsspeilet 2020. Retrieved from https://www.udir.no/tall-og-forskning/finn-forskning/tema/utdanningsspeilet-2020/del-1/videregaende/

Vitaz, M., & Poletanovic, M. (2020). Data-Driven Learning: The Serbian Case. *EL.LE*(2). doi:10.30687/ELLE/2280-6792/2019/02/009

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

Walker, C. L., & Shore, B. M. (2015). Understanding Classroom Roles in Inquiry Education: Linking Role Theory and Social Constructivism to the Concept of Role Diversification. *SAGE Open, 5*(4), 215824401560758. doi:10.1177/2158244015607584

Weimer, M. (2013). *Learner-centered teaching : five key changes to practice*. In.

Wicher, O. (2020). Data-Driven Learning in the Secondary Classroom: A critical evaluation from the perspective of of foreign language didactics. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. London: Routledge.

Dissertation articles

1

Corpus Literacy and Applications in Norwegian Upper Secondary Schools: Teacher and Learner Perspectives

Petter Hagen Karlsen and Marte Monsen, Inland Norway University of Applied Sciences[1]

Abstract

This study explores corpus literacy and pedagogical corpus applications among in-service teachers who have corpus linguistics in their educational background, along with their students of English in upper secondary school in Norway. In order to investigate the multifaceted nature of educational practice, the study took a wide perspective encompassing aspects of digitalization in education, general pedagogy, and corpus linguistics. The students (n=154) answered an online questionnaire and four teachers were interviewed. The data are discussed in light of a technology integration framework that takes into account teacher knowledge, resource availability, teacher beliefs, as well as student epistemologies. Findings show little to no corpus literacy among the students, with the exception of one student. The interviews show that the teachers have largely avoided corpora in their teaching practice. One reason why the teachers chose to do so was due to inaccessibility related to paywalls, registration requirements and functionality restrictions. Other potential reasons that were uncovered were tied to their perception of learner competencies, how the affordances of corpora fit their curricular focus, and how their epistemic beliefs about language learning influenced their choices.

Keywords: pedagogical corpus application; English language; corpus literacy

1. Introduction

Corpora may offer a range of pedagogically valuable affordances in the English as a foreign language [EFL] classroom. However, according to Cardona, Didriksen, and Gjesdal (2014: 1), they have not been used to a satisfactory degree in Norwegian schools. Pedagogically oriented corpus linguists claim that applying corpora in the classroom has several benefits. As a resource, corpora grant access to authentic language examples, as opposed to artificial, engineered examples and rules (Boulton & Cobb 2017; Leńko-Szymańska & Boulton 2015), contain information on context, frequency, collocations, and distribution (Boulton 2010), and provide opportunity for the examination of lexical and phraseological

[1] The authors have no competing interests to declare.

Karlsen, Petter Hagen and Marte Monsen. 2020. "Corpus Literacy and Applications in Norwegian Upper Secondary Schools: Teacher and Learner Perspectives." *Nordic Journal of English Studies* 19(1):118-148.

patterns, and language variation through register and genre information (Farr 2008). These features have several reported pedagogical affordances. For instance, examining corpora can raise the learners' language awareness by examining real language behavior through authentic language (Leńko-Szymańska & Boulton 2015: 3), it can open up for learners following their own interests in language discovery (Bernardini 2004), and it can promote the acquisition of new learning skills and learner autonomy by focusing on learner-centered corpus interactions with the learner being akin to a discoverer or researcher (see Boulton & Cobb 2017; Cheng, Warren & Xun-feng 2003; Johns 1991; Millar & Lehtinen 2008).

This study examines the role of linguistic corpora in four secondary schools, by interviewing four teachers on their corpus use and classroom experience and by investigating the corpus literacy of the students in seven of their classes by means of a questionnaire. These teachers all had a course in corpus linguistics as part of their teacher education. In an experiment with pre-service teachers, Leńko-Szymańska (2017: 234) concluded that a semester-long corpus course may not be sufficient to give future teachers the technological, corpus linguistic, and pedagogical skills required for the pedagogical application of corpora. Similarly, the four teachers participating in this study had a semester-long corpus linguistics course; however, this study examines in-service teachers, a perspective that has not been explored as much. One study in Norway of in-service teachers' familiarity with corpora found little to no familiarity with the concept among the respondents (Kavanagh in preparation). Our belief is that the perspectives from in-service teachers and their students can offer ecological explanations of why corpora are reportedly underused in English language education in Norway.

We argue that by not only exploring the teachers' perspectives, but also the learners', one gets a better understanding of the dynamics of the classroom, which may reveal opportunities or barriers for pedagogical corpus integration. In order to explore the corpus literacy of the students, we pose the research question: (1) How familiar are upper secondary school students with corpora? In light of the student data, we turn our focus to the teachers to see how their corpus education translated into practice, while also inquiring into their general teaching practice to make visible potential challenges and opportunities for the integration of corpora and corpus tools into the classroom. Thus, we pose a second research question: (2) What beliefs do teachers express about corpora as a pedagogical tool?

We turn first to theoretical perspectives associated with corpora as pedagogical tools. Next, since the research is seen through the lens of technology integration and teacher and student beliefs, we will be discussing these in some detail, before we turn to describing our method in Section 3. In Section 4.1, we first examine the degree to which corpora have affected the participants, before we investigate and discuss the reasons for the teachers' choices in Section 4.2. Ultimately, we conclude in Section 5.

1.1. Corpora as a pedagogical tool

In this study, we focus on the direct application of corpora in the classroom setting, i.e. when teachers or learners interact with them directly (Römer 2011). Such direct applications naturally encroach more noticeably on the everyday experience of both teachers and learners and on educational practice, which is why we mainly examine direct use of corpora. Classroom corpora integration can range from inductive scenarios, where the learners make generalizations from the corpus data, to deductive scenarios, where the learners check some lexico-grammatical rule or pattern against data (Liu & Lei 2017: 31-34). It can also range from learner-autonomous, learner-centered to teacher-led activities (Mukherjee 2006: 12). The most inductive, and learner-centered approach to the pedagogical application of corpora is Data-Driven Learning [DDL] as proposed by Johns (1991), where the learner is conceptualized as a researcher, the computer an informant, and the teacher a director or coordinator (Johns 1991). However, Gilquin and Granger (2010: 359) have defined DDL as 'using the tools and techniques of corpus linguistics for pedagogical purposes', thus broadening the concept considerably.In this study, we take a broad perspective on DDL, as defined by Gilquin and Granger (2010), since we are looking for any sign of direct corpus utilization. Moreover, Callies (2019: 247) defines corpus literacy as "a multicomponential set of complex skills".Drawing on Mukherjee (2004) and Dalton-Puffer (2014), Callies (2019: 247-248) presents four components of teachers' corpus literacy:

> 1. Understanding basic concepts in corpus linguistics: What is a corpus and what types of corpora are available and how? What can you do—and cannot do [sic]—with a corpus?

2. Searching corpora and analysing corpus data by means of corpus software tools, e.g. concordancers: What is corpus software and how can it be used to search a corpus? How can corpus output be analysed?
3. Interpreting corpus data: How may general trends in language use/change be extrapolated from corpus data?
4. Using corpus output to generate teaching material and activities: How can you make use of corpus material for teaching purposes?

Although this construct has teachers in mind, we adopt some of the points as categories of analysis against which we can examine students' knowledge of corpora. This adoption is possible, we would argue, because students would need a semblance of some of these skills in order to exploit corpus data independently and effectively. A couple of things should be noted, however. First, the points listed above suggest the need for declarative knowledge. We acknowledge that students' exposure to corpora may have been less explicit and less terminology-centered. According to Frankenberg-Garcia (2014), it is unnecessary to teach learners what corpora are or train them to perform linguistic analyses; they simply need to know how to look up their specific language quandaries. We therefore sought to ask questions that could elicit exposure to corpora that has not resulted in declarative knowledge. When we discuss students' familiarity with corpora, we mean in a broad sense any sign of student-corpora interaction. Second, the fourth point is clearly directed at teachers and is therefore not suitable to our current discussion. Third, we emphasize that (a) even if students report to have no knowledge of corpora, it does not necessarily mean that the teachers have avoided it completely, and (b) even if students report having comprehensive knowledge of corpora, it does not necessarily mean that the source of said knowledge is their current English teacher. We only treat these teachers as a likely source of potential corpus knowledge due to their educational background.

Several challenges have been outlined for applying corpora pedagogically. Some have argued the digital medium to be a possible challenge for corpus integration (e.g. Boulton 2010), while others suggest that today's computer access and digital competence among the younger generation may open the door for corpora (e.g. Cardona, Didriksen & Gjesdal 2014; Flowerdew 2009). The latter positions, however, can be tied to Prensky (2001: 1) generationally bounded term 'digital natives', which has been problematized by Bennett, Maton, and Kervin (2008: 783),

whose review of several studies on youths' digital competence concluded that there is little evidence for a distinct generation with sophisticated digital technology skills and learning preferences. Another issue is that many of the major corpora were created for linguistic, not pedagogical, purposes (Braun 2007: 308), and may not be suitable for younger learners. Lastly, Leńko-Szymańska (2014: 261) lists digital and computer skills, usability, access, cost, and lack of knowledge about corpora as other or related problem areas.

2. Theories of Language Learning in DDL

DDL has been linked to three distinct yet connected theories of language learning: the noticing hypothesis, constructivist learning, and sociocultural theory (Flowerdew 2015: 16-19). According to the noticing hypothesis, learners are more likely to acquire language competence if their attention is consciously directed toward linguistic features. Schmidt (2001, 2010) maintains that noticing is a prerequisite for understanding. He also suggests that conscious work to 'notice the gap' between the learners' own output and target language input is necessary to overcome errors in L2 language production (2010: 724). Related to the noticing hypothesis, and as we will argue also relevant in connection to DDL, are theories about linguistic and metalinguistic awareness, for example as put forward by Bialystok (2001). According to for example James and Garrett (1992), Bialystok (2001), Purpura (2004), and Van Essen (2008), the ability to evaluate and explain language practices (metalinguistic awareness) will support language learning, and in order to develop such awareness, students need to pay conscious attention to linguistic features. One could argue that DDL, for example focusing on investigating recurring features in the target language, enhances opportunities for students to focus and reflect on lexico-grammatical features in a way that fosters metalinguistic awareness.

Flowerdew (2015: 25) gives examples of studies that '[illustrate] how a corpus can be mediated to address learner needs in line with principles associated with constructivist learning', by providing opportunities for students to choose among ways of investigating linguistic features. For instance, by choosing or switching between formal grammar guides and corpus searches, the learner is a co-constructer of knowledge about language at different levels.

Flowerdew (2015: 27) also maintains that DDL provides opportunities for peer-to-peer interaction with a view to mentoring and feedback. In this way, fellow students may be able to scaffold each other's learning. According to Vygotskyan sociocultural theory, learning occurs when individuals interact with others within their zone of proximal development (Lightbown & Spada 2013: 47). Vygotsky defines the zone of proximal development as 'the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in cooperation with more capable peers' (Vygotsky 1978: 86). In contrast to e.g. processability theory, the emphasis of the model of proximal development is on social interaction itself rather than on the internal processing of information (Cook 2016: 249). According to Vygotskyan theory, learners gain control over their cognitive processes through mediation, and knowledge is internalized when they engage in social activity (ibid.). As Flowerdew (2015: 27-28) shows, certain ways of working with DDL foster learning in a Vygotskyan sense, such as by facilitating peer-to-peer interactions during groupwork where the students also engage with the corpus. In the context of this study, a primary interest is in how students and teachers conceptualize corpus-based approaches, and how it is strongly related to their views of how knowledge production and learning occur.

2.1. DDL as technology integration

Working with corpora in the classroom requires of the teacher technical, pedagogical and corpus linguistics skills (Leńko-Szymańska 2017; Mukherjee 2006). Similar categorization is made in the technological pedagogical content knowledge [TPACK] model, which examines 'the complex web of relationships between users, technologies, practices, and tools' (Koehler & Mishra 2005: 132). The model's predictive power has been questioned due to the lack of evidence for the discreetness of each domain (e.g. Archambault & Barnett 2010; Graham 2011). However, it is beyond the scope of this article to clear up this knowledge construct, and we do not seek to track, test, or explain the interrelations between the teachers' knowledge(s). Instead, we treat knowledge of technology, corpus linguistics, and pedagogy as separate concepts only to operationalize the research, and for descriptive purposes. By technology,

we are particularly interested in the teachers' experience with digital technologies and their influence on classroom practice. In the case of corpus linguistics knowledge, we refer back to Mukherjee's (2006) construct corpus literacy as a basic knowledge of corpora. Finally, pedagogy involves the teachers' choices and reflections on teaching practices, learning, planning, and teacher-student interactions.

However, these abovementioned knowledge domains are not the only aspects said to influence successful integration of technology in the classroom. In the wider framework of technology integration, of which we count pedagogical applications of corpora to be part, teacher knowledge is but one element. According to Kim et al. (2013: 77), teacher knowledge, together with environmental readiness, make up a first-order barrier to technology integration. Environmental readiness can be thought of as the access to and availability of necessary resources and conditions to make use of a novel technology. For instance, integrating corpora in the classroom requires the access to a corpus that is free or not too expensive with an easy-to-use interface, as well as time to plan out its application and implementation, and access to computers or tablets.

Moreover, teacher beliefs pose a second-order barrier to technology integration (Kim et al. 2013: 77). Borg (2011: 371) describes teachers' beliefs as "propositions individuals consider to be true and which are often tacit, have a strong evaluative and affective component, provide a basis for action, and are resistant to change". Teachers' beliefs are theorized to 'facilitate or hinder practice by serving to filter, frame, and guide experience, decisions, and actions' (Fives & Buehl 2012), and thus serve as a potential gatekeeping mechanism. There is no stable, unified definition of the teacher beliefs construct (Pajares 1992; Skott 2015). To operationalize this wide concept, Pajares (1992: 316) suggests discussing educational beliefs about…, e.g. about epistemology, about student performance, etc., which can be thought of as the distinct but related subsystems of one's whole belief system (Fives & Buehl 2012). One can be conscious or unaware of these beliefs, and they can vary in stability. However, beliefs show a degree of plasticity to experience and interaction (Fives & Buehl 2012: 473-475).

In this study, we take a particular interest in the teachers' thinking about the nature of knowledge, about pedagogy, and about their students. We do not seek to track the origins of these beliefs (e.g. Levin & He 2008), nor do we discuss how to transform them (e.g. Fives & Buehl 2014); we

only examine them as possible explanatory terms for the teachers' corpus-related choices. We argue that teachers' beliefs may reveal (dis)congruence between teacher epistemologies and the theoretical underpinnings of DDL. In addition, we also investigate student beliefs—what some researchers refer to as student epistemologies—as potentially influential for corpus use, as we believe student influence to be central to classroom practice. In line with Elby's (2009: 139) broad conceptualization of personal epistemologies, we take students' epistemologies to mean their views about the nature of knowledge and knowing, and about learning.

3. Methods and Materials

3.1. Participants

Several teachers were asked to participate in the study with the selection criterion that they taught English in secondary school in Norway. The teachers who were ultimately selected and agreed to participate were four teachers of upper secondary school, three with a Master's degree in English language education, one with a Master's underway, and all with some formal corpus-related training. Specifically, they all had a similar, semester-long corpus course at Master's level, which focused on both corpus analysis and the critical evaluation of corpus-based studies on the one hand, and the ability to facilitate corpus-based classroom activities and to use corpora as an aid for independent learning on the other. Thus, their courses had both a teaching and learning component, and a corpus linguistics component. They all taught at different schools. Since this selection of participants was done based on the somewhat specific characteristic that they were English teachers with some corpus background, the sampling strategy can be defined as purposive sampling, where '[…] researchers handpick the cases to be included in the sample on the basis of their judgement of their typicality or possession of the particular characteristic(s) being sought' (Cohen, Manion & Morrison 2018: 218). We therefore make no claims about the wider population, but try instead to access individuals with particular knowledge and experience.

The students participating in the survey were those from the teachers' classes. The study includes both general studies and vocational studies programs, but the number of participants is greatly skewed toward general studies, as is the case in the Norwegian upper secondary school system.

Table 1 provides an overview of the schools, the teachers, and the number of participating students in each of the teachers' classes (Class 1, 2 & 3). As shown in the 'percentage of total' column, the number of participating students per teacher is skewed. In the case of Marcus' classes, his second and third class were of a particular vocation that tend to have smaller class sizes than general studies classes. Moreover, not all of their classes were available for participation.

Table 1. Overview of schools, teachers, grades and number of students who completed the questionnaire (Class 1, 2, 3). The teachers' names are pseudonyms.

School	Teacher	Class 1	Class 2	Class 3	**Total**	**Percentage of total**
W	Nora	16	21	-	37	24%
X	John	29	30	-	59	39%
Y	Marcus	20	5	6	31	20%
Z	Sarah	8	19	-	27	17%
Total					**154**	**100%**

The project was approved by the Norwegian Centre for Research Data [NSD] prior to the commencement of the research, including processes of gathering informants, giving information and receiving consent, in order to ensure ethical conduct that guarantees the participants' anonymity in any subsequent publications. A consent letter with detailed description of the research were signed by everyone involved and all participants were informed about their option to withdraw consent at any point without consequences.

3.2. Research design

In this study, we sought both student and teacher perspectives to examine the presence of corpora in classroom practices, and to uncover potentials and challenges for successful corpus implementation. By reaching out to as many parties as possible within the different classes, we sought to obtain democratic validity, which is concerned with whether or not the research is done in collaboration with everyone who has a stake in the research (Newton & Burgess 2008: 26). The students answered an online

questionnaire. Subsequently, the teachers were interviewed based partially on preliminary questionnaire findings.

As discussed in Section 2.1, the domains of technology, pedagogy, and corpus linguistics are important aspects of corpus classroom integration (Leńko-Szymańska 2017; Mukherjee 2006). These constructs formed an overarching framework for both the student questionnaire and teacher interviews. Thus, the following categories emerged: (1) technology-integration, including digital technologies such as computers and websites, (2) pedagogy, relating to general pedagogical approaches, preferences, and language teaching and learning theories as discussed by Flowerdew (2015), and (3) corpus literacy, as described by Mukherjee (2006), concerning familiarity and competence with corpora and corpus linguistics.

The questionnaire was constructed using the program Checkbox (Checkbox.com), and by following recommendations from Cohen, Manion and Morrison (2018) about question length, question order, and item type. It was piloted by a group of students from a different school than the ones involved in the research proper. The questionnaire contained thirty-seven items which were predominantly Likert-type items, but with open and multiple-choice questions interspersed. The students could choose either English or Norwegian versions of the questionnaire. All students chose Norwegian with one exception. The responses to the open-ended questions were translated by the first author. The questionnaire took around twenty minutes to complete, was distributed by the teachers, and was completed in class.

The interviews were semi-structured and sought to ascertain the teachers' perspectives on the use of techniques related to corpora or DDL, and to uncover their perceived challenges and possibilities associated with these concepts. The interview guide contained twenty-three questions starting from why they chose to be English teachers and concluding with specific corpus-related questions. The interviews were conducted in Norwegian, audio recorded, transcribed by one of the researchers, and presented as excerpts translated into English by the same researcher who transcribed them. Each interview lasted a little over an hour, with the exception of Marcus', which lasted almost two hours due to his frequent elaborations. Interviews, transcriptions, and translations were all conducted by the same researcher, the main author of this paper.

3.3. Data analyses

Cohen, Manion and Morrison (2018: 842) suggest frequencies and cross-tabulations for non-parametric data, which is the case for the ordinal and nominal data collected through our questionnaire. Since the sample was neither random nor sufficiently large, no effort was made to generalize statistically to a larger population. The data are presented in frequency tables, as percentages, or illustrated through bar charts. The interview data were subject to a qualitative analysis process of segmenting, coding and reassembling (Boeije 2010: 77-79). The data were first segmented based on emerging themes, and given labels—or codes—to mark these themes. These new units were analyzed both in relation to each other and in light of the questionnaire data, which made up the process of reassembly (see Boeije 2010: 76), a process that relies on the subjective discretion of the researchers.

4. Results & Discussion

In the following we present and discuss the findings of the study. In Section 4.1, we explore whether or not the students have any corpus experience by looking for signs of corpus literacy (as presented in Section 1.1) in the data. We also examine the teacher interviews to see whether or not they have implemented corpora in their practice. Note that students' declarative knowledge of corpora or lack thereof does not necessarily reflect on the teachers' practice but could also be signs of previous experiences, indirect corpus work, or even forgetfulness. Section 4.2 presents and discusses the findings relating to why corpora have not featured to any notable degree in the classes involved. It is subdivided into three parts, namely the corpus dimension, the digital dimension, and the pedagogic dimension. Each of these parts corresponds to Leńko-Szymańska's (2017) claim that teachers need technical, pedagogical, and corpus linguistics skills to integrate corpora in their teaching (see Section 2.1). These categories are not treated as discreet units but are instead used as categories to structure the discussion. Lastly, we also discuss some possibilities for integrating corpora as a language-learning tool for upper secondary students and make some tentative suggestions for ways forward.

4.1. Corpus literacy and application

4.1.1. Students' corpus experience

Figure 1 shows the digital resources the students report using in school based on the open-ended question 'Which digital tools and/or webpages do you use at school?' The item was open-ended to see if any corpus-related tool was mentioned. None of the specified webpages or dictionaries was corpus-related. Keep in mind that these responses might reflect their immediate associations, rather than being exhaustive. Note that the categories shown in Figure 1 are made by the researcher based on students' free responses. For instance, general answers like 'PC' or 'Mac' were categorized as 'hardware'.

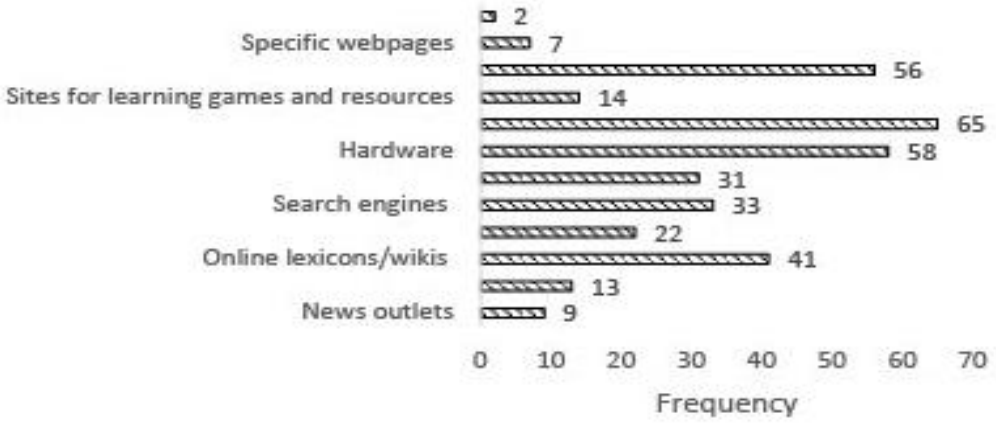


Figure 1. Digital tools the students report using with frequency information

When asked directly in the questionnaire, 94% of the students reported no familiarity with corpora. The remaining 6% who claimed familiarity with corpora were presented with an open-ended question to specify their experience. As can be seen in Table 2, only two of the eight students who had heard of corpora, both taught by Marcus, attempted a description. Student 6 gave a short, concise definition, while student 7 gave a more ambiguous answer. Student 2's answer 'nri' is most likely a typo of the Norwegian word for 'no' (nei).

Table 2. An overview of students' answers to what they have heard about corpora, coupled with each students' teacher

#	Student responses	Teacher
1	Someone has mentioned it, don't know what it is.	Nora
2	Nri	Nora
3	Something about cooperation.	John
4	Have only heard the word before, do not remember the context.	John
5	From our teacher Marcus.	Marcus
6	That it is a tool to find words and expressions that are used in natural contexts among English speaking people. That it is how the language is used in everyday speech by those who have English as their mother tongue.	Marcus
7	It has been mentioned. It helps you find different words for things.	Marcus
8	The teacher took some of us aside and went through what corpus was and whether it was a topic we wanted to learn more about later.	Marcus

Finally, some students responded that they were familiar with a selection of corpus tools in a multiple-choice question (Figure 2). The category 'Other' gave them the opportunity to elaborate freely on what other tools they might have heard of. None of the subsequent responses was corpus-related.

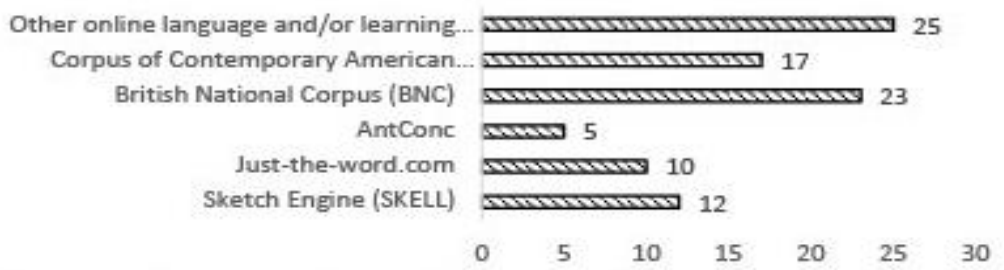


Figure 2. Student answers to the item 'Have you heard about any of the online tools listed below?' with frequency information. The item was multiple-choice, allowing students to select any number of the alternatives given in the chart

4.1.2. The teachers' implementation of corpora

Having examined the students' corpus experiences, we interviewed their English teachers to investigate how corpora had affected their practice. We considered these corpus-trained teachers as the most likely source of the students' experience with corpora. When asked how corpora had influenced her teaching practice, Sarah responded:

> (1) Way too little… in an ideal world I would have gladly used it a lot, but the students… they utilizing it themselves is a way too extensive process. I think the interface of these sites are far too difficult and troublesome… it takes too much time to learn [how to utilize] corpora, and I do not think [the students] are in a place language-wise where they would have any benefit from it, either.

She emphasized that corpora were too far removed from the student everyday experience, as they have multiple subjects at the same time, and may not possess the linguistic interest one might wish. When Nora was asked how corpora had influenced her practice, she answered:

> (2) I told my students about a corpus assignment, or a corpus study, I did myself … and told them about the concept, but we have not used it in teaching, we have not looked into it. I have not implemented it in my work yet. In a way, I have the project that I am trying to complete with my Master's thesis, and then there is my job on the other hand. As of now, there is a thin, thin, thin common thread between the two.

Elaborating on the time she told her students about her own research—a study of discourse markers in learner English—she stated,

> (3) It was really just a digression, since we were talking about formal and informal language, slang and such, so I suppose it had its source in etymology and language, but they were politely interested [laughter].

John also explained how corpora had not influenced his teaching practice in any significant way, but he thought it had made him more conscious about choosing dictionaries, (4) 'if you search for a word [in a corpus] you get it in use—in actual use—as it is used someplace, as opposed to a invented, made-up example. I find that useful, and I think the students would too, if they can understand what it's all about, that is.'

Notably, the number of students who answered affirmatively on some of the tools seen in Figure 2 does not match the number who claimed familiarity with the term 'corpora'. The teachers found this peculiar, with the exception of Marcus who clarified that he had once taken those he

described as his brightest students aside and introduced them to the very basics of corpora. Since it was a multiple-choice question, the teachers speculated that, although unlikely, perhaps their students had encountered corpora in lower secondary, or, more likely, that they had answered because of an association with the name or initialism (e.g. BNC), rather than skipping the question. Judging from the teachers' skepticism, the item type, and the other student responses about corpora, questions arise about the item's reliability. It is likely that the students responded through some association with the words or acronyms, rather than familiarity with the concept.

These results shed some light on the research question 'How familiar are upper secondary school students with corpora?' None of the students brings up any corpus-associated tools when asked about their digital habits in general, and when asked directly, only a small minority claims any recognition of the terms 'corpus' or 'corpora' in an educational context. Furthermore, those who report familiarity with the phenomenon mostly supply unsubstantial description of the terms, with the exception of two students. Cardona, Didriksen and Gjesdal's (2014: 6) observation that corpora have had little impact in Norwegian secondary schools seems to be reflected in this group despite their English teachers' formal corpus training.

Although the students' reported experiences do not necessarily reflect what the teachers have actually done in the classroom, these findings suggest that corpora have not been featured enough to make an impact on their declarative knowledge, which might further serve as evidence for Leńko-Szymańska's (2017: 272) conclusion, as presented in Section 1.1, that a one-semester corpus course may not be sufficient to make pre-service teachers confident in using corpora pedagogically in their own practice, albeit indirectly. A central premise of corpus literacy, i.e. familiarity with corpora and their uses, is not met by any of the students save one. It is interesting how, despite having discussed corpora with a small group of students, only one of Marcus's students could produce a meaningful definition of corpora. On the whole, however, their lack of knowledge is unsurprising, as the teachers predominantly opted away from using corpora in their practice. A few challenges to DDL are revealed in Sarah's comments about her students' language level, their lukewarm language interest, and the poor user-interfaces of corpora. Next, we further

investigate the reasons behind the teachers' corpus-related choices in their educational practice.

4.2. Why not corpora?

4.2.1. The corpora dimension

Three of the teachers perceived corpora as an academic tool. According to Marcus,

> (5) [Integrating corpora in my feedback system] is shown not to be possible, because the corpus world is made one hundred percent for academics, so it becomes sort of 'Do you want access? That will be 3000 dollars a year'. It is not doable for a teacher.

Similarly, John stated that his own corpus education seemed more useful for academic pursuits than for applications in schools. Sarah explained, (6) 'I think that, on a high academic level, it is probably valuable, but [in order to apply it] in schools by having the students use it themselves, for example, it has to be made simpler and more accessible'.

The issue of accessibility is echoed by Nora, as can be seen in this interaction:

> (7) Interviewer: How accessible are corpora and corpus resources?
>
> Nora: Inaccessible, are they not? . . . I have experienced it as a threshold, but not extremely so.
> I have not been lying awake at night because of it [laughter].
>
> Interviewer: But would it have been worse if every student needed access?
>
> Nora: Yes. If every student had to apply for access, it would be a hassle, and I would lean toward 'no'.

Marcus was the only teacher who explicitly tried to implement corpora in his practice. He has developed a system where he, with the click of a button, added pre-written comments when giving feedback electronically. For instance, if student's writing was unidiomatic, he could press a button that added a comment in the text explaining idiomaticity and providing examples. For language feedback, he attaches language examples from learner dictionaries. He commented that he attempted to link the example function to corpora, but the corpora had paywalls, and no Application Programming Interface (API) that allowed for his needs.

Two interesting perspectives can be gleaned from these statements. First, the accessibility to corpora is an environmental readiness barrier (i.e. inaccessibility, costliness, etc.), as discussed in Section 2.1, since many of the larger corpora have paywalls and are too expensive to introduce to an entire class. It is also a question of teacher knowledge, as they seemed to have mainly interacted with larger corpora such as COCA and the BNC, while being unaware of smaller corpus resources without restrictions such as cost, registration or search limit. On the one hand, this barrier can be explained through the teachers' lack of time, patience, or familiarity to orient themselves in the corpora landscape. On the other hand, it can be seen as a limitation of their education, which seems to have involved a narrow selection of larger, academic corpora.

Second, the teachers' descriptions of corpora as academic research tools reflect Braun's (2007: 308) observation that many of the larger, accessible corpora are designed for linguistics, not pedagogy. If the teachers conceptualize corpora as best suited for the academic level, it is not surprising that they also find it too daunting for their own students. For instance, if their academic experiences with corpora made great demands on their lexico-grammatical knowledge, it can explain why Sarah sees her students' lacking language interest and competence as probable obstacles should they navigate a corpus themselves (see Section 4.1). This belief, which likely is a product of the teachers' own corpus experience, can act as a deterrent for classroom corpus-integration.

4.2.2. The digital dimension

As highlighted in Section 1.1, the digital aspect of corpora has been discussed as a source of both obstacles and possibilities to DDL (e.g. Boulton 2010; Cardona, Didriksen & Gjesdal 2014). Table 3 shows that the majority of the students agree to varying degrees with the statement that they are used to working with digital tools in class (92%). The majority also preferred working with digital tools (77%) as opposed to working without them, with fifteen percent neither agreeing nor disagreeing. Seemingly, the students' beliefs about digital technologies are positively skewed, with mobile phones excluded due to some of the teachers banning them in the classroom. They report familiarity with digital technology in educational contexts, and, more interestingly, they seem to see a relationship between learning and digitalization, but not

quite as strong as the relationship between finding information and digital tools. These beliefs might suggest that they view learning, which can be related to their personal epistemologies (Elby 2009), as positively affected by digitalization and the increased access to information. Their exposure to digital technology appears to be unquestionable and their impressions of digital resources in education are largely positive, so in that sense they are digital natives. There is, however, a clear discrepancy between student beliefs and teacher beliefs, which becomes apparent in the teacher interviews.

The teachers were asked to comment on the results from two of the items in Table 3, namely the students' self-perceived ability to learn new computer programs and digital tools, and their perception that these tools have made learning easier. Nora commented on the relationship between learning and digitalization:

> (8) I think it has created winners and losers… it is only a personal, anecdotal hypothesis I have, but I think perhaps the gap between those who are real digital natives and those who do not have a clue is getting wider and wider. I think some of the students have frighteningly bad technological skills… It may well be that they are good with social media, and that they are really good at reading blogs—I don't actually know—but they have difficulty finding documents stored on their Macs, for example, or uploading things to [the school's online learner platform] Canvas.

Sarah described her students' digital skills as (9) 'surprisingly bad, despite having grown up with the Internet'. She further elaborated that they are uncritical of information and will often ask her whether a source is good. She, like Nora, pointed out that they lacked good search strategies: (10) 'When they search for information, they ask questions [to the Google search engine], like sentence-length questions… with question marks and everything.' John calls his students digital competency worryingly low and says that they are quick to ask before really trying themselves. Marcus describes his students' digital competency as 'very, very low', despite having grown up in today's society. He claims they lack fundamental skills, such as cloud-storing, changing languages in Microsoft Word, keyword searches, and awareness of which program to use when opening files.

Table 3. Upper secondary students' responses to their familiarity and preferences concerning digital tools in their education. N=154. All students responded to all items: 100%=154 respondents.

Items	Strongly agree	Partially agree	Neither agree nor disagree	Partially disagree	Strongly disagree
I am used to working with digital tools in English lessons (computer, tablet, the internet).	64 %	28 %	6 %	2 %	0 %
I am used to working with my mobile phone in English lessons.	6 %	7 %	16 %	32 %	39 %
I find it easy to learn new computer programs and digital tools.	37 %	38 %	17 %	7 %	2 %
I learn more in classes where I can use a computer, a tablet or other digital tools.	29 %	35 %	24 %	10 %	1 %
I prefer working with digital aids/tools, as opposed to working without them.	42 %	35 %	15 %	7 %	1 %
I think technology and digital tools have made it easier to find information.	76 %	18 %	6 %	1 %	0 %
I think technology and digital tools have made it easier to learn new things.	49 %	41 %	8 %	2 %	0 %
I wish we would use computers, tablets or mobile phones more often in English class.	16 %	27 %	44 %	10 %	2 %

There is a clear discrepancy between the teachers' negatively skewed impressions of their students' digital competence, and the students' positively skewed impressions of their own abilities. These divergent impressions indicate a potential barrier to classroom corpus-integration. Positive teacher beliefs about technology have been shown to coincide with successful technology implementation in educational practice (Kim et al. 2013).If we keep in mind both the teachers' view of corpora as an academic tool and the lack of user-friendliness of most major corpora, it is unsurprising that the teachers are reluctant to introduce corpora in their teaching repertoire, given their classroom experiences. The technical

aspect of corpora has also been suggested as an obstacle for classroom corpus application; for instance, Farr (2008: 39-40) found that the biggest issues teacher training students had with corpus-based instruction, in which the students themselves utilized corpora, were technology and software related. Learners' frustrations with the technical aspects of corpora have been reported by other researchers as well (e.g. Boulton 2010). Given that both students and teachers in this study express confidence in using digital technology, it may seem the interface and usability of corpora are the greater issue, not technophobia as some suggest (see Boulton 2010: 539). Furthermore, several non-corpus-related applications (apps) are developed with language learning in mind that have user-friendly interfaces and usability metrics directly targeted at younger language learners. Most corpora do not offer the same advantages, as they are often developed for linguists with complex interfaces cluttered with search options and contain technical terminology and codes. There might therefore be a need to design corpus resources with simpler interfaces and fewer options to avoid confusion.

In addition, the teachers said they were surprised by their students' low digital competency, which they expected to be higher for someone growing up in the digital age. This may indicate that the 'digital natives' assumption (see Section 1.1) was part of the teachers' beliefs at the onset of their careers but has since been challenged. As Nora expressed, the digital age has created 'winners and losers', a divide between the technologically competent and the strugglers. Corpora as pedagogical resources introduce not only novel ways of studying language, but also novel ways of utilizing digital technology. In addition, it may potentially widen the gap between the 'winners and losers' of the digital age by building on the varied and unsteady foundation of assumed student digital competency, and consequently alienate the learners who already find simpler digital work difficult.

4.2.3. The pedagogical dimension

Next, we turn to views of pedagogy and how it might explain why the teachers have largely opted away from utilizing corpora in their everyday practice. First, we examine the teachers' descriptions of their English lessons in general. Second, we try to discuss the teachers' epistemological beliefs. These beliefs were explored partly through the teachers'

reflections on students' answers to the questionnaire items 'I think the teacher should be able to answer any question I have about English during English class' (see Figure 3) and 'I do not think the teacher should ask me questions during English class that he/she does not have the answer to himself/herself' (see Figure 4). Both parts endeavored to discover how the teachers' general pedagogical and epistemological beliefs could potentially influence the integration of corpora in teaching.

All four teachers described their lessons as topic-focused (English-speaking countries, global issues, self-chosen topics, etc.), with each topic often spanning several weeks. When inquired about their language-specific teaching, the teachers stated that they had no or limited explicit grammar focus. Marcus taught vocabulary explicitly with his vocational classes through profession-specific word lists, by having the students practice these words in their vocational environment, and through Quizlet (Quizlet Inc. 2019), a website that lets you create flashcards and quizzes for students. The explicit vocabulary teaching for his general studies classes were mainly through feedback on student writing via his system of pre-written comments and attached examples of use described above (see Section 4.2.1). John said his explicit language-focus was through feedback and color-coding of errors. Nora's vocabulary teaching was implicit through working with different topics. If someone needed additional tasks during a lesson, she might send them to Exploringenglish.cappelendamm.no (2012), a site offering English language learning resources to students and teachers, to work on a specific grammatical phenomenon. When discussing corpora and their uses, however, the teachers saw them as tools for etymological research, as reference works for grammar, collocations, and prepositions connected to idiomatic language-use, or as tools for comparative linguistic analyses.

If the teachers' approaches to pedagogy are seen in light of how they described corpora in certain capacities, we start seeing some possible obstacles to corpus integration. Firstly, none of these capacities is directly related to the study of social and cultural themes, as appears to be the pedagogical and curricular focus in these classrooms, but rather directly tied to linguistic inquires. Thus, one could argue that the teachers' beliefs about the usefulness of corpora in education are colored by their experience of corpora as tools for explicit language research, and not as tools for social, cultural or pragmatic study. Their focus on social and cultural themes, which will naturally be influenced by curriculum

demands and number of available teaching hours, does not open for time-consuming, explicit language learning activities such as inductive DDL activities.

Secondly, there seems to be a discrepancy between the teachers' approaches to language study—i.e. through corrective feedback or implicit acquisition—and the explicit language focus theorized in DDL—e.g. inductive learning and the noticing hypothesis (see Section 2). This difference shows a potential mismatch between the teachers' epistemic beliefs, and the type of language study suggested in DDL. In other words, the teachers approach language learning as a case of immersion and feedback, whereas the constructivist approach of DDL puts language at the forefront, favors salience, and requires longer sessions of interpretive work with language data—as its connection to the noticing hypothesis also indicates (see Section 2), DDL requires conscious attention to language. If these disparities in epistemic beliefs are a reality, there are fundamentally different epistemic foundations between DDL and the language pedagogy of the teachers. These differences are a second-order barrier that is hard to grasp, as beliefs about language learning are likely to be central to a language teacher's belief system. We must therefore be sensitive to not only the teachers' resources, knowledge, and needs, but also to their deeply held epistemological beliefs.

On the one hand, we should heed the advice that, '[…] any incremental step or change should be sensitive to the current needs of teachers' (Kim et al. 2013: 83), and we might need to rethink the strong inductive drive in DDL. On the other hand, we should not be afraid to challenge teachers' entrenched epistemological assumptions in manners that may transform their pedagogy to serve student needs in different ways. To achieve this, we suggest scholars interested in pedagogical applications of corpora work closely with teachers and students to examine possible ways to add corpus to their pedagogy incrementally, avoiding invasiveness, and with a mind to complimenting already established practice.

To further investigate the epistemic beliefs of the teachers and students, we included the questionnaire items 'I think the teacher should be able to answer any question I have about English during English class' (see Figure 3), and 'I do not think the teacher should ask me questions during English class that he/she does not have the answer to himself/herself' (see Figure 4.) in the student questionnaire, and asked the teachers to comment on the results. We deemed this information

interesting as DDL is suggested to entail a shift in teacher style from instructor and authority to coordinator and director (Johns 1991: 3). Both questions open up for interpretations from the respondents, but they sought to elicit their views on teachers as experts.

Although mostly skewed toward agreement, both graphs show a high number of students who take no position (33 and 36 respondents respectively). These results may be due to the items' ambiguity, as they can be interpreted in several ways. This ambiguity was pointed out by the teachers when they were asked to comment on the result, which in turn questions the internal validity of the items. It should not be read as an answer to how students regard their teachers as experts, but rather as an indication that most of the students regard the teacher as an knowledge authority in some capacity (frequencies of 95 and 76 respectively)—in what capacity would require more in-depth dialogue with the students.

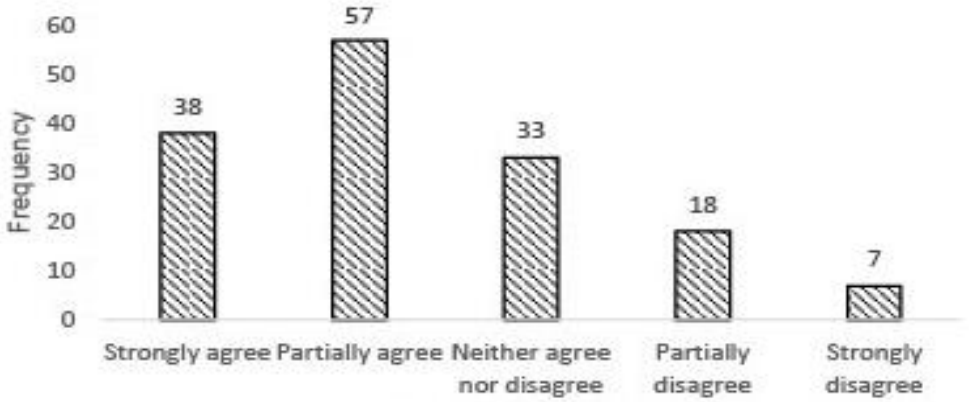


Figure 3. Students' answer to the questionnaire item: 'I think the teacher should be able to answer any question I have about English during English class' by frequency.

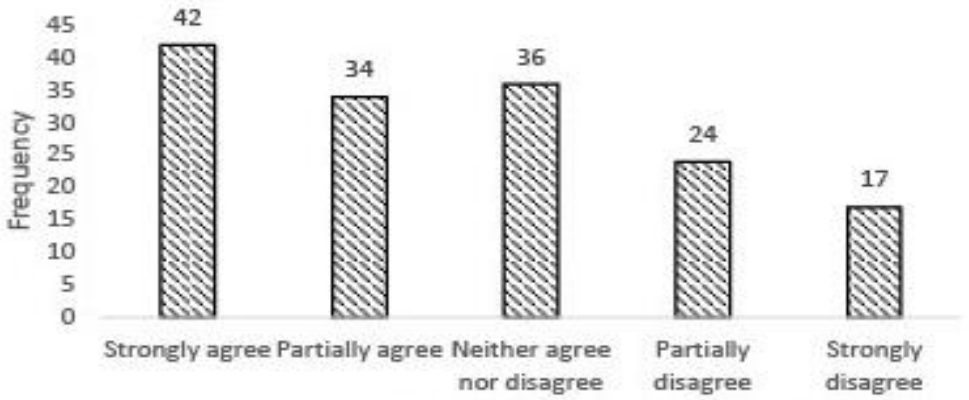


Figure 4. Students' answer to the questionnaire item: 'I do not think the teacher should ask me questions during English class that he/she does not have the answer to himself/herself', by frequency.

Rather than presenting these results in isolation, we used them as prompts in the teacher interviews. The teachers were asked to comment on the graphs in Figures 3 and 4 and give their perspectives on teachers as knowledge experts. Both Nora and John expressed that being a knowledge expert is difficult in the modern information society. Marcus said that a teacher should be someone who knows enough, but still dares to acknowledge when s/he does not have the answer. He did emphasize that a teacher should know the basics, such as concord, and should avoid [L1] interference mistakes. Sarah avoided closed questions with her pupils, and instead focused on the big questions, analyses, critique, and reflection. Nora also mentioned reflection-based schools as more time-appropriate, (11) 'if there is a grammar task, a multiple-choice task, or a sort of right-or-wrong task, yes [there could be a key], but that's a bit "out", isn't it? I am thinking that schools have become more reflection-based'. Similar divisions between language and social topics were made by the other teachers. Sarah said she had more expectations toward her linguistic knowledge, as had her students:

> (12) [The students] ask about something there and then, because they have discovered something that does not coincide [with what they know], or they find a pattern that is different, and I am not always able to say 'it is like this because of this' in a simple way without using subject terminology.

There appears to be a division between the teachers' beliefs and expectations of linguistic knowledge (grammar, orthography, etc.) versus knowledge of social and cultural topics (e.g. the #MeToo-movement). They largely approach the English subject similarly, with students using internet resources to explore topics, and they acknowledge the impossibility of being an expert in every such topic. They also seem to promote open exploration, and openness to varying student answers. However, they appear to have a stronger expert identity tied to the linguistic side of their profession, albeit to varying degrees.

Arguably, DDL in its most inductive form promotes a student-centered and explorative approach to language learning not unlike the pedagogies promoted by the teachers when working with varying topics. Here, the teacher's expert, authoritative role shifts toward a facilitative one where language teaching is no longer the teaching of proscriptive grammar rules, and learners become constructors of their own language models (see Section 2). Our findings suggest that the teachers appear more skeptical toward affording their students the same freedom in discovering language, as they are with social and cultural topics, although Sarah's observation that students come to her with unusual language patterns they have discovered seems like prime territory to involve DDL. This skepticism can be interpreted as the teachers' beliefs about their students' language competence, it can be due to their non-linguistic pedagogical focus and priorities, or it can be indicative of their epistemic beliefs about language, i.e. they view language knowledge and thematic socio-cultural knowledge as different domains requiring different pedagogies. Whether one or more of these interpretations are accurate is beyond the scope of the current study to conclude on; however, they provide some potential perspectives for future inquiries.

5. Concluding Remarks

As our findings suggest, the teachers avoided corpora in their teaching practice, and the students reported little to no experience with corpora. First-order barriers, such as paywalls, registration requirements, and functionality restrictions (e.g. APIs), appear to have dissuaded them from integrating corpora in their teaching practice. On one hand, this issue concerns the question of whether corpus resources are available, pedagogically appropriate, and freely accessible. Most corpora are made

for linguistic inquiry in mind, not pedagogic use (Braun 2007), and there are few corpora that are both freely available and appropriate for younger learners. On the other hand, the issue points to the teachers' knowledge about what kinds of corpus resources are out there, as they seem to have been exposed mainly to large, general corpora such as the BNC and COCA.

There are two implications of these findings. First, there is a dearth of free, user-friendly, pedagogically available corpora, which means that new materials need to be developed if corpora are to be of interest to pre-tertiary language teachers. A solution lies in material designers' willingness to create such resources based on teacher and student feedback, and corpus scholars' research on pedagogical corpus use (e.g. Pérez-Paredes 2019; Wicher 2019). Second, the corpus courses offered during teacher training should focus not only on large, general corpora, but also on what resources and possibilities are out there in terms of pedagogically appropriate corpora and how to explore them. This point is further emphasized by the teachers' description of corpora as fundamentally academic and not immediately relevant to their students, a view that is likely linked to their own experiences with corpora. If corpus resources are to find their way into the pre-tertiary classroom, teacher training should provide the type of resources appropriate to the teachers and their students' needs.

Perhaps more difficult to remedy are the discrepancies between teachers' negative view and students' positive view on student digital competency, and the way the teachers see linguistic knowledge in a different way than they see social and cultural knowledge. In case of the former, suggestions have been made to remove the digital component by relying on paper-based concordances (see Boulton 2010); however, the students reported a clear preference toward utilizing computers in their learning process. One direction could be to better understand user-friendliness from a learner perspective when designing corpus resources, so that the burden of explaining and understanding cluttered or advanced interfaces does not fall solely on the teacher. It could also imply that giving students free rein when exploring corpora can lead to confusion, and that the teacher might have to offer instructions. In case of the latter, the way teachers perceive knowledge is difficult to change, if change is even desirable. Instead, future research could explore affordances of corpora

beyond analyzing frequency data and concordance lines and look at how corpora can fit in an educational environment that is topic centered.

This study has revealed some potential barriers to using corpora directly in the classroom by seeking the perspectives of both teachers and learners and by examining classroom practice more holistically. It should be noted that the study only looked at four teachers and that the analyses were predominantly qualitative, which makes generalization difficult. The research was also limited to the upper secondary context and teachers with a common corpus background. Further research is therefore needed for primary and lower secondary school, and with teachers from different educational backgrounds. Although our findings suggest little direct use of corpora, most of the teachers were ultimately positive about trying corpora in their teaching practice. We believe that corpora have a place in the EFL classroom, if we can be more sensitive to teacher and learner needs, and help co-design corpora, corpus resources, and corpus approaches that will benefit and complement the educational process beyond 'what works'.

References

Archambault, Leanna M., & Joshua H. Barnett. 2010. Revisiting technological pedagogical content knowledge: Exploring the TPACK framework. *Computers & Education, 55*(4), 1656-1662. doi:10.1016/j.compedu.2010.07.009

Bennett, Sue, Karl Maton, & Lisa Kervin. 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

Bernardini, Silvia. 2004. Corpora in the classroom: An overview and some reflections on future development. In *How to use corpora in language teaching*, edited by John M. Sinclair (pp. 15-36). Philadelphia: J. Benjamins.

Bialystok, Ellen. 2001. *Bilingualism in development: Language, literacy and cognition. Cambridge*. Cambridge, UK: Cambridge University Press.

Boeije, Hennie. 2010. *Analysis in qualitative research*. Los Angeles: SAGE.

Borg, Simon. 2011. The impact of in-service teacher education on language teachers' beliefs. *System, 39*(3), 370-380.

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

———, & Tom Cobb. 2017. Corpus use in language learning: A meta-analysis. *Language Learning, 67*(2), 348-393.

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

Callies, Marcus. 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. vi, 267 pages). Amsterdam; Philadelphia: John Benjamins Publishing Company.

Cardona, Margrete D., Anders A. Didriksen, & Anje M. Gjesdal. 2014. Korpusbasert undervisning i fremmedspråkene: La elevenes nysgjerrighet sette dagsorden. *Acta Didactica Norge, 8*(2), 1-26.

Checkbox.com. 2018. Checkbox (Version 2018Q2). Massachusetts, USA: Checkbox Survey Inc. Retrieved from https://www.checkbox.com/

Cheng, Winnie, Martin Warren, & Xu Xun-feng. 2003. The language learner as language researcher: putting corpus linguistics on the timetable. *System, 31*(2), 173-186. doi:http://dx.doi.org/10.1016/S0346-251X(03)00019-8

Cohen, Louis, Lawrence Manion, & Keith Morrison. 2018. *Research methods in education* (Eighth edition. ed.). London; New York: Routledge.

Cook, Vivian. 2016. *Second language learning and language teaching* (Fifth Edition. ed.). New York: Routledge.

Elby, Andrew. 2009. Defining personal epistemology: A response to Hofer & Pintrich (1997) and Sandoval (2005). *Journal of the Learning Sciences, 18*(1), 138-149.

Exploringenglish.cappelendamm.no. 2012. Exploring English. Retrieved from https://exploringenglish.cappelendamm.no/

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:10.2167/la414.0

Fives, Helenrose, & Michelle M. Buehl. 2012. Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us? In *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors.* (pp. 471-499). Washington, DC, US: American Psychological Association.

Flowerdew, John. 2009. Corpora in language teaching. In *The handbook of language teaching*, edited by Michael H. Long & Catherine Doughty (pp. 327-350). Chichester, U.K.; Malden, MA: Wiley-Blackwell.

Flowerdew, Lynne. 2015. Data-driven learning and language learning theories: Whither the twain shall meet. In *Multiple affordances of language corpora for data*-driven, edited by A. Lénko-Szymánska & A. Boulton (pp. 15-36). Amsterdam/Philadelphia: John Benjamins publishing company.

Frankenberg-Garcia, Ana. 2014. How language learners can benefit from corpora, or not. *Recherches en didactique des langues et des cultures [En ligne], 11*(1), 1-8.

Gilquin, Gaëtanelle, & Sylviane Granger. 2010. How can data-driven learning be used in language teaching. In Anne O'Keeffe & Michael McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (1st ed., pp. 359-370). London; New York, NY: Routledge.

Graham, Charles R. 2011. Theoretical considerations for understanding technological pedagogical content knowledge (TPACK). *Computers & Education, 57*(3), 1953-1960. doi:10.1016/ j.compedu.2011.04.010

James, Carl, & Peter Garrett. 1992. *Language awareness in the classroom*. London; New York: Longman.

Johns, Tim. 1991. "Should you be persuaded": Two samples of data-driven learning materials. *Classroom Concordancing ELR Journal, 4*, 1-16.

Kavanagh, Barry. (In preparation). *Bridging the gap from the other side: The uses English teachers have found for corpus methods.* Manuscript submitted for publication.

Kim, Chanmin, Min K. Kim, Chiajung Lee, J. Michael Spector, & Karen DeMeester. 2013. Teacher beliefs and technology integration.

Teaching and Teacher Education, 29, 76-85. doi:10.1016/j.tate.2012.08.005

Koehler, Matthew J., & Punya Mishra. 2005. What happens when teachers design educationaltechnology? The development of technological pedagogical content knowledge. *Journal of Educational Computing Research, 32*(2), 131-152.

Leńko-Szymańska, Agnieszka. 2017. Training teachers in data-driven learning: Tackling the challenge. *Language Learning & Technology, 21*(3), 217-214.

———. 2014. Is this enough? A qualitative evaluation of the effectiveness of a teacher-training course on the use of corpora in language education. *26*(2), 260-278. doi:10.1017/S095834401400010X

———, & Alex Boulton. 2015. Introduction: Data-driven learning in language pedagogy. In *Multiple affordances of language corpora for data-driven learning*, edited by Agnieszka Lénko-Szymánska & Alex Boulton (pp. 1-14). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Lightbown, Patsy, & Nina. M. Spada. 2013. *How languages are learned* (Fourth edition. ed.). Oxford: Oxford University Press.

Liu, Dilin., & Lei Lei. 2017. *Using Corpora for language learning and teaching*. Alexandria, VA: TESOL International Association.

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

Mukherjee, Joybrato. 2006. Corpus linguistics and language pedagogy: The state of the art – and beyond. In *Corpus technology and language pedagogy: new resources, new tools, new methods*, edited by Sabine. Braun, Kurt Kohn, & Joybrato Mukherjee (pp. 5-24). New York: P. Lang.

Newton, Paul, & David Burgess. 2008. Exploring Types of Educational Action Research: Implications for Research Validity. *International Journal of Qualitative Methods, 7*(4), 18-30.

Pajares, M. Frank. 1992. Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307-332. doi:10.3102/00346543062003307

Pérez-Paredes, P. 2019. The Pedagogical 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. London: Routledge.

Prensky, Marc. 2001. Digital natives, digital immigrants. *On the Horizon, 9*(5), 1-6.

Purpura, James E. 2004. *Assessing grammar*. Cambridge, U.K. ; New York: Cambridge University Press.

Quizlet Inc. 2019. Quizlet: Learning tools and flashcards, for free. Retrieved from https://quizlet.com/

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

Skott, Jeppe. 2015. The promises, problems, and prospects of research on teachers' beliefs. In *International handbook of research on teachers beliefs* (Vol. [13]), edited by Helenrose Fives & Michele G. Gill. New York: Routledge.

Van Essen, Arthur. 2008. Language awareness and knowledge about language. In *Knowledge about language* (Vol. 6), edited byJasone Cenoz & Nancy H. Hornberger. New York: Springer.

Vygotsky, Lev. S. 1978. *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wicher, Oliver. 2019. Data-Driven Learning in the Secondary Classroom: A critical evaluation from the perspective of of foreign language didactics. In P. Crosthwaite (Ed.), *Data-driven learning for the next generation: Corpora and DDL for pre-tertiary learners*. London: Routledge.

2

3

Nordic Journal of Language Teaching and Learning

Educational Roles in Corpus-Based Education: From Shift to Diversification

Petter Hagen Karlsen
Inland Norway University of Applied Sciences

1. Introduction

There is a growing interest in using linguistic corpora for pedagogical purposes, judging by the upward trend of empirical studies on the subject the last two decades (see Boulton & Cobb, 2017, p. 363). Applying "[…] the tools and techniques of corpus linguistics for pedagogical purposes" has become known as 'data-driven learning' [DDL] (Gilquin & Granger, 2010, p. 359), a term that gained impetus through Johns (1991), who envisioned the learner as a researcher, the teacher as a director/coordinator of learner-initiated research, and the computer as an informant (pp.1-3). Thus, DDL is the direct applications of corpora in the classroom by teachers and/or learners (Leech, 1997; Römer, 2011). Many corpus scholars argue strongly for using corpora in the classroom, citing benefits such as access to authentic language and increased language awareness (Boulton & Cobb, 2017; Lénko-Szymánska & Boulton, 2015), student-centered discovery learning facilitating students' own language interests (Bernardini, 2004), and student autonomy and new learning skills (Boulton & Cobb, 2017; Cheng, Warren, & Xun-feng, 2003; Johns, 1991). DDL promises new types of data, new skills, and new directions for the learner. However, early DDL literature provides hands-on activities, but lacks descriptions of the broader picture of teachers' decision-making (Wicher, 2020, p. 31). Instead, the teacher's role is repeatedly challenged, while focus is put on the student. Boulton (2009) states, "It may even be, in some cases, that learning is more effective without a teacher, i.e. when learners discover things for themselves" (p. 37), while Gilquin and Granger (2010) suggest, "DDL […] requires that the teachers take risks, and agree to 'let go' and let the student take pride of place in the classroom" (p. 367). These propositions are made despite the fact that most instructors in DDL research are DDL scholars and not regular teachers (Vyatkina, 2016, p. 207), and that there are few qualitative pre-tertiary studies (Pérez-Paredes, 2019). The teacher's role is de-emphasized, and the students are pushed to the forefront; yet, these roles are often ill-defined. What is more, proponents of DDL make assumptions about the secondary school classroom that become problematic when trying to realize a mostly university-level tested approach with younger learners. This study problematizes some of these assumptions in light of student experiences and feedback and adds perspectives from inquiry-based education and student-centered teaching to establish a more nuanced understanding of educational roles in DDL.

Section 2 looks at the conceptualizations of teachers and learners in DDL and reviews theoretical perspectives on educational roles. Section 3 covers the research design, participants, data collection methods, and analysis. The interview data are presented in Section 4 and discussed in Section 5. Lastly, Section 6 concludes the paper by looking at implications and limitations of the research, and at future directions.

2. Educational Roles in Corpus-Based Education in Light of Pedagogic Theories and Previous Research

In addition to providing new learning activities and resources, DDL seeks to reconceptualize educational roles in terms of student-centered teaching and learning. The relationship between corpus-based education and pedagogic principles has been explored in recent literature; Meunier (2020) argues for constructive alignment in DDL, asking whether curriculum, teaching methods, and assessment tasks are "consistently and coherently aligned" (p. 13), while Wicher (2020) discusses DDL's fit with Task-Based Language Teaching. Following this trend, the current section draws on two pedagogic approaches similar to DDL – learner-centered teaching and inquiry-based education – as they offer better developed concepts of role taking and role acquisition.

2.1 The Teacher and the Learner in DDL

DDL is predominantly conceptualized as learner centered (e.g. McEnery & Xiao, 2011), as the approach "usurps the traditional roles of the teacher/researcher and student […]" with the teacher no longer a language expert responsible for teaching and research, but a facilitator, while the student becomes both investigator and learner (Cheng et al., 2003, p. 175). Ideas of lessened teacher control or even involvement appear to permeate the discourse (cf. Boulton, 2009; Gilquin & Granger, 2010 in Section 1) and can be traced back to early conceptions of DDL where Johns (1991) suggested that one could "cut out the middleman as far as possible [i.e., the teacher] and […] give students direct access to the data" (p. 30). The focus is firmly put on the student. For instance, Bernardini (2004) emphasized DDL's discovery learning aspect and student interests with the learner-as-traveler metaphor wherein students actively explore corpora in open-ended ways with the teacher no longer a limitless knowledge source (pp. 22-23). Similarly, Papaioannou, Mattheoudakis, and Agathopoulou (2020) reference the teacher-as-facilitator premise, stating: "the teacher is no longer the source or expert of all knowledge [but should provide] guidance and support when necessary" (p. 187), and emphasizing student discovery and responsibility over a passive recipient role: "The learner is an agent who investigates language, finds clues about the meaning and relations between linguistic items, and generate [sic] hypotheses regarding meaning and rules" (ibid.). In a study by Gatto (2020), students used the Web as a corpus while solving tasks, while the teachers acted as 'mediators' and 'supervisors' following a short introduction on quantitative evidence in language studies (pp. 114-115). Unfortunately, the mediator and supervisor descriptors were not elaborated on. Breyer (2009) argues that the teacher role has long been neglected in the DDL literature. She created a pre-service teacher course where teachers acted as both learners using and analyzing corpora, and teachers designing and assessing corpus-based learning materials, which lead to the teacher becoming a learner, guide, material designer, and mediator for novice learners (pp. 154-167). Thus, Breyer (2009) showed the many teacher responsibilities and requirements in a DDL setting and began to problematize vanishing teacher role in the DDL literature. Furthermore, it is argued that teachers need a basic corpus literacy in order to successfully integrate corpora in their classrooms, which entails being able to search, interpret, analyze, generalize, and extrapolate from the corpus data, as well as design teaching materials (Callies, 2019). Students would face many of the same requirements in becoming independent corpus users (see Lee, Warschauer, & Lee, 2020, p. 346). What is more, it is argued that teaching with corpora requires digital/technological, content, and pedagogic knowledge (Leńko-Szymańska, 2017; Meunier, 2020). Although these claims outline the requirements and expectations put upon both teachers and learners, they lack descriptions of how these principles are engendered and enacted in the process of acquiring and preforming different roles in the DDL classroom, an issue that becomes increasingly problematic when the discussion is moved to pre-tertiary education and based on assumptions about how roles are

developed and enacted in the pre-tertiary classroom. In Section 2.2, two pedagogical perspectives are presented that can inform and diversify the conceptualization of roles in DDL.

2.2 Perspectives from Learner-Centered Teaching and Inquiry-Based Education

DDL, inquiry-based approaches, and student-centered teaching share many features, and among these are a constructivist foundation, learner-centeredness, and the use of researcher-emulated processes by the learners. These perspectives offer a more nuanced framework of educational roles, from which DDL can benefit.

Learner-centered teaching "[…] shifts the role of the teachers from givers of information to facilitators in student learning" (Darsih, 2018, p. 33). According to Weimer (2013), this shift includes increased student discovery, problem-solving, and peer- and self-assessment, while the teacher does less lecturing intended to transfer information and instead prioritizes preparation, positive learning climates, student evaluation, and concrete modeling of the learning process (pp. 72-84). Doyle (2011) argues that the teacher-as-facilitator role means "providing [students with] an environment for engagement; a set of resources such as questions, articles, research findings, problems, and/or cases to engage with; and using assessment tools that provide the learning with meaningful feedback" (p. 52), all the while encouraging full participation, mutual understanding, and shared responsibility (p. 53). These principles provide concrete aspects of the teacher's roles in learner-centered classrooms.

Moreover, inquiry-based learning involves observing, posing questions, reviewing evidence, investigating, predicting, and using tools to gather, analyze, and interpret data (The National Research Council, 1996, p. 23) and "learning through question asking based on curiosity and interest" (Walker & Shore, 2015, p. 3), all of which align well with the principles of DDL. Taking on the inquiry role as a teacher means mastering these skills and fostering them in one's students, who in turn must adopt them. Walker and Shore (2015) propose a four-stage process of role acquisition for both students and teachers (pp. 7-10). The first stage is *exploration*, where students and teachers familiarize themselves with the expectations of an inquiry-based classroom, such as students' initiative taking, teamwork and creativity. After exploration comes *engagement*, where students "formally adopt and engage in an inquiry student role" (Walker & Shore, 2015, p. 8). In this stage, students can create questions, take initiative, discuss, organize information, and interpret data independently and collaboratively (ibid.). The third stage is *stabilization*, in which the student or teacher are committed to the changes that come with the approach. Here, conflicts felt due to traditional role habits have been resolved, and teachers and learners "[…] positively value collaboration, are comfortable with problems that are not well defined, look for patterns across knowledge areas, think imaginatively and critically, and acknowledge multiple solutions to problems" (Shore et al., 2009, quoted in Walker & Shore, 2015, p. 9). In the final stage, *diversification*, students and teachers "[adopt] additional and varied roles within the classroom" (Walker & Shore, 2015, p. 9). For instance, students can take the role of teacher, team leader, hypothesizer, presenter, audience or explorer (ibid.). Clearly, reaching the final stage entails great versatility and flexibility from all participants to enable them to take on different roles when it is required and/or fruitful. The concepts of stabilization and diversification could be what is required to reach a normalization of corpus application in schools, as discussed by Chambers (2019), i.e. when corpus consultation has become second nature to the language learners.

The shortcomings in the conceptualization of roles acquisition and enactment in DDL, as outlined in subsection 2.1, may thus be alleviated by principles from inquiry and learner-

centered teaching. Examining DDL in light of these principles and students' actual opinions and experiences also enables us to highlight and problematize some of the assumptions in the DDL literature about the secondary school classroom. The following two research questions are posed:

1. How do DDL proponents' assumptions about the upper secondary classroom and its educational roles align with the experiences and opinions of students?
2. How can perspectives from inquiry-based education and student-centered teaching inform the conceptualization to educational roles in DDL?

3. Methods and Materials

The present study draws on student interview data from a case study wherein multimodal corpora designed for pedagogic purposes were integrated into two first-year upper secondary school classes in collaboration with their regular English teacher. As shown in Table 1, the interviews were group-based, with five students per group and two groups per class. Two classes, one with 33 students and one with 36 students aged fifteen and sixteen, participated in the research. The interviews were semi-structured and thus encouraged digressions, discussions and elaborations. The differences between interview times reflect the students' willingness to engage and elaborate. The interviews were audio and video recorded, conducted and transcribed in Norwegian, and then translated by the author/researcher for the purposes of presentation in subsequent publications. Part of the interview data, as well as observational data from the corpus-integration period, are reported on elsewhere (see Karlsen in preparation). While Karlsen (in preparation) focuses on the students' and teacher's engagement with and experience of the BACKBONE corpora, the present study draws on other parts of the interview data to explore students' opinions on education roles.

Table 1. *Composition of interview groups in relation to class and group affiliation, group sizes, gender balance, and time spent on the interview.*

Class	Group	Female	Male	Total students	Interview times
A	A1	3	2	5	51 minutes
A	A2	2	3	5	48 minutes
B	B1	3	2	5	34 minutes
B	B2	2	3	5	38 minutes
Total		10	10	20	

The corpora were available on the BACKBONE webpage and consisted of videotaped interviews and searchable transcripts featuring speakers of English varieties traditionally neglected in classrooms (e.g., Irish-English). These corpora were designed with pedagogical purposes and younger learners in mind (Kohn, Hoffstaedter, & Widmann, 2009). This resource was chosen due to the lack of other appropriate, freely available resources at the time of the research implementation in the autumn of 2019, although new multimedia corpus resources are currently in development (see Hirata, 2020). The corpus-integration period lasted two weeks and was implemented by the abovementioned teacher, but planned and designed by the author/researcher. Students worked with tasks that gave instructions on corpus exploration, as well as open-ended writing assignments resembling their mid-term exams. The more structured tasks provided step-by-step instructions on how to navigate the corpus tool and required the students to explore, compare and discuss different aspects of the website's functions (e.g., frequency lists, concordance lines or interview excepts) and a varieties of topics (e.g., idiomatic expressions). The tasks and lesson plans can be accessed here as online appendices. Since the

tasks and lesson plans were designed by the author/researcher, restrictions were put on the freedom of the teacher. After discussions with the teacher prior to the integration period, the following organizational outlay of the classroom was agreed on: student groups of six arranged around their desks. How the teacher introduced, approached or concluded activities was not specified; however, the tasks were designed in a structured manner that guided the students through how to the use of the BACKBONE website in order to reduce the degree to which the teacher had to familiarize himself with the BACKBONE website.

The data were analyzed through a process of segmenting, coding, and reassembling (Boeije, 2010). The chronologically gathered data were divided into segments based on codes and reassembled thematically for comparison. Some codes were established before the integration period based on theoretical constructs (concept-driven coding), but more codes were added during coding, transcriptions, and interpretations (data-driven coding). This dual coding approach was chosen because the semi-structured interviews produced data that could not be predicted beforehand.

The study and data collection were approved by the Norwegian Center for Research Data [NSD] (The Norwegian Directorate for Education and Training, 2019). Informed, free consent was given by everyone involved. Anyone could withdraw at any point without consequences. Everyone involved received a letter detailing the study with a high degree of openness.

4. Presentation of Data

The interviews show students' perceptions and expressed preferences regarding predominantly student-centered work with minimal instructions from the teacher. Note that the teacher was not given any pointers as to how he should act in the classroom, only task-sets that detailed how the students could approach the corpus resources. Citations from the interviews are sequentially numbered as indicated by the numbers in parenthesis. Student speakers are named S1, 2… 5, the interviewer is denoted as I, and the interviewed groups are denoted as either A or B for class and 1 or 2 for group in line with Table 1 (see Section 3).

Each group was first asked what their English lessons usually looked like. All four groups mentioned self-study as a dominant feature of the lessons. B1 was most adamant about it, as shown in the following interaction:

(1)
S1: There's a lot of self-study, almost exclusively.
S2: Yes.
S3: Mhm.
S2: I think it can be difficult because he [the teacher] has an idea of what we're supposed to do, but then you do something you thought you should do. There is little lecturing, and it's sometimes difficult knowing what to do.

B2 experienced the lessons as varied depending on the topic, with some group work, some PowerPoint, and some self-study. They described a scenario in which the teacher provided them with articles online they had to read in order to write a text. A1 described having experienced a lot of teacher-centered situations where the teacher introduced a new topic for the whole class for anything between five and twenty minutes before they work on their own. Similarly, A2 agreed there was a lot of self-study but with a PowerPoint introduction. One A2 student did not appear to see the self-study as a positive, as he commented on the decrease in student-centered lessons in the following way:

(2)
I: So, you feel there is a high degree of student-centered lessons where you're in control?

S1: Yes, it has become better since we began [upper secondary], but there was a lot in the beginning.

The students were asked during the interviews to comment on how they thought the teacher had functioned during the corpus-implementation period. All groups voiced that they would have wanted more of an introduction before being 'thrown into' the tasks. The teacher's introduction was a presentation of the researcher/observer and a discussion between the students and the teacher about what a recipe was, how to follow one, and an explanation of how the task-sets were to be followed like recipes. One student in A1 commented:

(3): S: "I thought the task was really difficult, but I'm not sure how much our teacher knew either. I don't know if the teacher knew much about the project or the website [the corpora]. I don't know if he had used it much beforehand."

Similar concerns were raised in A2:

(4) I: What about the teacher? How did you experience him this period?
S1: I didn't notice him much. I felt we were just working on the computer and that was it.
S2: […] often when I asked him he didn't seem to have a reason as to why and [for] what we were using [the corpus tool], and it was demotivating to continue and not understand. All he could do was point us to the next step of the list [the task-sets], but not help us actually understand it.
S1: I sort of felt that we had to understand everything ourselves, had to learn it by ourselves.

In addition to wanting a more thorough explanation of *what* to do with the task-sets, the students of B1 agreed they also wanted a more thorough explanation of *why* they had to do it.

(5) I: How did the teacher function during this period?
S1: I think it might have been difficult to be the teacher since there was so much joking around and groups so large that no one could concentrate or pay attention, and he couldn't help everyone at once. When all the groups are sitting around wondering what to do, it might be a hint that he should have explained a bit better at the beginning, because all the groups sat around not knowing what to do.
I: Was it explained eventually, or was it just confusion and him trying to put out fires?
S2: He tried to explain eventually, but it came a bit late.
I: What could he have done differently?
S1: I mean, he doesn't know what the problem is when he starts, but perhaps try to pick up on it earlier and not when we're mid-lesson or almost done.

As with A2, B2 did not notice the teacher much during the period except for when they raised a hand to get help. One student commented:

(6): S: "We could raise a hand and ask for help, as we're used to. I think it would've been harder if he had stood before us and explained, because then I wouldn't have followed it at all because it is much easier if you manage to ask questions yourself, because if everyone is listening to him at the same time, he cannot help everyone simultaneously."

Something both A-groups and B1 wanted was an introduction on how to write 'discussing texts' with research questions. This request was not only linked to the corpus period, but to their experience from the semester as a whole, which entailed these types of writing assignments. The conversation was trigged by the writing assignment they were given as part of the corpus-based period. In A1, students wanted an example text, a formula on how to write discussing texts like a math formula, and a list of words to use/not use. A2 had struggled to formulate research questions to their texts that also took the use of corpora into account. The students in B1 wanted to be taught more about text structure in general and have more clearly defined teacher expectations about what constitutes a good text. They missed initial training on

how to write, structure and word such texts, especially what words not to use. One student wanted a blackboard demonstration, and had the following conversation with a peer student:

(7) S1: I feel like a lot of the basics disappear.
S2: Yes. It would help us knowing how to [write a text], so that we maybe could focus on writing English […]

B2 did not comment on writing training, but instead wanted an assignment with a self-chosen topic, as opposed to the two topics given.

(8) I: How did you experience writing a bit more research-based tasks where you have to use several sources and create a research question?
S1: It depends a little on the topic. If it's a topic you find interesting, it's not that difficult. That makes finding information and websites easier, but it becomes harder once you're not that interested in the assignment.

The organization of the classroom was a point of discussion that revealed variation among the interviewed groups. While all of the groups expressed that they preferred some degree of variation, four of the students in A1 favored working alone with the option of asking their teacher or peers. Working alone forced them to solve tasks themselves, instead of copying other – often stronger students – which made them feel like they got less out of the lesson. One student said pair-work was all right, while another thought the big groups – as during the corpus-based period – made concentration difficult. Conversely, B1 felt too often left to themselves and that there were too few teacher-led lectures. Although they wanted a mixture of both, student-centered lessons dominated. This organization had led to one student feeling as if he did not always live up to the teacher's expectations, and another felt that the teacher had too high expectations of their abilities. The A2 students did not elaborate much beyond stating they wanted variation, that too big groups made concentration difficult, and that too much of the same became boring. Lastly, B2 had the following discussion:

(9) I: So do you prefer working in groups as opposed to working alone?
S1: It varies, but ideally, yes.
S2: It depends a lot on the task though. It's easier to be alone if you're writing a text.
S3: Yes.
I: I can imagine. Or else you'll become distracted?
S2: Yes.
S3: Yes.
I: How do you like situations where the teacher talks a lot and you pay attention or only participate a bit?
S4: It becomes a bit boring. You quickly lose concentration because you grow tired of listening all the time.
S1: But from time to time it can be somewhat pleasant that you don't have to do anything yourself and just listen, if you're focused.
I: Just lean back and…
S1: And in a way try to receive the information, but often when he is telling [us something] it is useful information, which makes for a lot of writing in addition to listening, and that makes it difficult to follow.

As shown in this section, the students express many diverse opinions and preferences regarding their educational experience. In the following section, these perspectives are discussed in relation to DDL proponents' assumptions about secondary education and in light of inquiry-based and learner-centered approaches.

5. The Emergence and Diversification of Educational Roles in Corpus-Based Education

In the following section, the interview data are discussed in relation to perspectives on educational roles in DDL and related didactic perspectives (see Section 2) to answer what educational roles are afforded by corpus-based lessons, and what student opinions reveal about these educational roles. The interview excerpts are referred to by their number in parenthesis. In section 5.1, the students' preferences and previous experiences are discussed against assumptions about changes DDL would bring to the classroom. Section 5.2 discusses the students' feedback from the implementation period and what it shows about roles in light of the abovementioned theoretical perspectives.

5.1 Students' Previous Experiences and Assumptions about Roles

The first point that should be addressed is the students' impressions of their English lessons prior to researcher interference and how they align with the positive changes suggested in DDL. Note that these are retrospective generalizations prone to biases; nevertheless, their input gives an impression of their previous classroom experiences. Based on student utterances, there is much to suggest that the teacher already occupied a partly facilitative role. The students described their lessons prior to researcher interference as self-study lessons with some variation in group organization, and often prefaced by introductions on the topic from the teacher. The teacher would provide resources to explore, such as links to articles and problems to solve, and impart students with great responsibility for their own learning process (see Section 4). Although many of these aspects coincide with Doyle's (2011; see Section 2) notion of learner-centered teaching, the teacher's initial introductions fall closer to a teacher-as-teller/instructor role. In a sense, the teacher is therefore already demonstrating a degree of mobility between different roles. This mobility is in line with Darsih's (2018; see Section 2.2) observation that a *shift* of the teacher role from giver-of-information to facilitator may be less of a permanent change and more of an example of dynamic role taking based on teacher judgement.

Several challenges for learner-centered education emerged during the interviews. Firstly, excerpts (1) and (2) show that the students did not discuss learner-centeredness in exclusively positive terms, but experienced confusion and abandonment. They wanted more training and instruction on text writing, which reveals preferences toward a clearer teacher presence and concrete input on certain topics. Secondly, excerpt (7) shows that too much self-driven discovery while learning how to write texts was experienced as interfering with the students' language learning. The discovery aspect of corpus-based lessons appears to be largely concerned with discovery of language (see Section 2) but leaves the question of what other topics or elements of the learning process should be discovery-based and adhere to the abovementioned role shifts. Thirdly, it highlighted a potential challenge for the teacher to gage the degree of freedom and responsibility that should be entrusted to the students overall. This issue is expressed in one student's experience of not living up to the teacher's expectations, suggesting that increased student responsibility produces unclear expectations. Fourthly, the student discussion in (9) is a good example of the necessity of lesson variation; where one student found lecturing boring, another occasionally found it pleasant. The juxtaposition of the active DDL learner against the passive recipient, with the latter being construed as demotivated (Papaioannou et al., 2020; Section 2), may hold some value when discussing learner motivation; however, judging from this study, monotony due to lack of variation seemed a potential de-motivator and challenge to differentiation when transitioning completely to a learner-centered approach. Consequently, the call for teachers to "'let go' and let the student take pride of place in the classroom" (Gilquin & Granger, 2010; see Section 1) raises three concerns. Firstly, that this shift has partially occurred and DDL scholars are erecting a strawman against which DDL is compared. Arguably, successful corpus integration must rely on updated information about secondary classroom practices (for instance the fact that many

teachers in their everyday practice already alternate between several roles such as 'giver-of-information' or 'facilitator'), so that corpus-based techniques can be appropriately adapted to complement said practices. Secondly, to what extent should teachers relent control? Should everything be left up to discovery? If not, where should the line be drawn? Lastly, student-centered teaching appears to be held up as the de facto 'best practice' for learners and learning. This assumption should be examined more closely in relation to learning styles and preferences before it is acknowledged as an exclusively positive change to the classroom.

It should be stressed that the indications that there are already student-centered, inquiry-like practices is positive for corpus-based educational research. Viewed in terms of inquiry-based education (Walker & Shore, 2015; Section 2.2), parts of both the exploration stage and engagement stage could already be in place, such as teamwork, discussion, and creation of questions, offering corpus-based approaches a foundation to build on. Consequently, the role changes may not be as big a leap as first presumed since the secondary school teacher is not akin to a university lecturer. Simultaneously, through Walker and Shore's (2015) framework of role acquisition (Section 2.2), one can identify areas where students struggle with inquiry-based education – e.g. initiative taking or data interpretation – in order to focus on these elements in future research and foster them in corpus-based lessons. Notably, some students had the impression that the teacher was not sufficiently familiar with the project or the corpora (excerpts 3 and 4) and one student found it difficult to identify the teacher's expectations. These challenges suggest that the exploration stage (cf. Section 2.2) was not properly covered, i.e., "where students and teachers familiarize themselves with the expectations of an inquiry-based classroom" (Walker & Shore, 2015, p. 8). The data fit Walker and Shore's (2015) framework quite well, as such conflicts related to traditional role habits are predicted and must be resolved in order to reach stabilization (cf. Section 2.2).

5.2 Role Taking and Development during a DDL approach

Certain issues emerged from the students' feedback that touched on the facilitative teacher role and the investigative student role. During the two weeks of corpus integration, the students experienced the teacher as absent and unengaging (excerpts 3 and 4), and too slow to notice challenges that arose underway (excerpt 5). Based on their utterances, the students expected the teacher to frame the lesson, i.e. *why* are they working with the corpus; specify the usefulness of the tool, i.e. *what* can it be used for; motivate them; and instruct them on *how* it should be used once they got stuck, or even before the demand for aid arose (excerpts 4 and 5). As shown in (3) and (4), the students' opinions of the teacher as passive made them question his corpus expertise. From a research perspective, this feedback highlights issues of researchers functioning as material designers and teachers as implementers, as teachers get less ownership and a diminished overview over the lessons. At the same time, it reveals the subtleties of role taking and problematizes rhetoric such as that 'in some cases, […] learning [may be] more effective without a teacher' (Boulton, 2009; Section 1), or that students' curiosity and interests concerning language study are presumed to be realistic motivational forces (Bernardini, 2004). Additionally, it goes beyond the assumption that the problem lies in teachers' fear of giving up control (see Breyer, 2009). Rather, it shows how role taking becomes a negotiation of involvement, role definitions, and responsibility. Each student needs to invest in her/his new role and get involved, while the teacher needs to find his/her degree and type of involvement as a facilitator or instructor or coordinator, etc. The degree and type of involvement both needed and *preferred* require considerations that make clear-cut definitions of roles difficult. Lastly, as pointed out by Doyle (2011; Section 2), there must be a shared responsibility for the learning process among teachers and students. This responsibility is likely well understood by teachers,

but demotivated or uninterested students might not see it the same way. The teacher's job is to make space for and promote this shared responsibility, a difficult task if the students feel they are given too much or all of the responsibility, and are unaware of the preparation, organization and evaluation that go into student-centered teaching. According to Doyle (2011), the facilitative role entails providing an environment for engagement, resources, and evaluation (cf. Section 2), all of which were provided by the teacher in the current study, albeit to varying degrees. However, what was missing from the period of corpus implementation is Weimer's (2013; Section 2) notion of *modeling* in which teachers solve a task as a demonstration with meta-commentary. Added to the packed facilitative role suggested in DDL literature such as mediator, coordinator, and resource provider, the teacher partly takes the role of a learner and partly as an instructor. This technique is not student-centered, but teacher-led. It shows the usefulness of mobility between roles when teachers become learners and instructors, and students become observers and co-learners.

The argument here is to move from a discourse of the permanence of *role shift* toward the flux of *role diversification*. By understanding the different roles participants must take on based on the situation, we can get to a place where the advice for corpus-wielding teachers is diverse and specific. Work is already being done in this direction through mapping the required competencies and literacies of a corpus-wielding teacher (e.g. Callies, 2019; Lénko-Szymánska, 2017) or through incremental approaches to student-corpus interactions (e.g. Kennedy & Miceli, 2016). These studies could form a foundation for understanding the teacher-as-facilitator concept in a corpus-based environment. This basis alongside didactic theory can help us move past hyperbolic statements about the teacher no longer being an expert or source of knowledge (cf. Sections 1 and 2) and instead focus on where the teacher's expertise can fit, where it is still useful and necessary, and what types of new expertise are required for the teacher-as-facilitator to foster the student-as-investigator/researcher role as well as a multitude of other roles.

6. Limitations, Implications and Future Concerns

There are several limiting factors to this study. Firstly, it only follows one teacher and two classes and is restricted to the idiosyncrasies of this particular context. Second, the tasks were not typical concordance searches, but several exploratory activities. Third, the study was short, and a longer period would afford more space for an incremental approach. Fourth, the teacher, although involved in the planning, was very busy. Research where the teacher is even more invested would make the setting more natural.

Nevertheless, this study is a starting point for understanding the complexity of role taking in a corpus-based classroom. While student-active education is undoubtedly a positive development, one must be careful not to throw the baby out with the bathwater. There may still be value in some instruction, some information might be more easily shared with the whole class, and not all students respond similarly to student-centered discovery approaches. The teacher can be a facilitator, a director, a planner, a material designer, a learner, and even an instructor. Likewise, the learner can be an observer, constructor, detective, teacher, and even a passive recipient of some information. The language of 'role shift' should be traded for one of role taking and diversification, in which all participants "[adopt] additional and varied roles within the classroom" (Walker & Shore, 2015; see Section 2.2). In future research, more emphasis should be placed on discovering what roles are appropriate at the different stages of student corpus consultation. What sort of tasks should the teacher create? When is it appropriate to instruct students on the functionalities of a corpus? When should they be given more control

and how can the process be differentiated? These are just a few key questions that need answering for educational roles to be realistic and diverse in corpus-based education. Research in naturalistic classroom settings based on in-depth observations and participant experiences will offer a more in-depth understanding of the complexity of role taking so that we can reject the pursuit of *role shifts* and embrace the goal of *role diversification*.

References:

Bernardini, S. (2004). Corpora in the classroom: An overview and some reflections on future development. In J. M. Sinclair (Ed.), *How to use corpora in language teaching* (pp. 15-36). Philadelphia: J. Benjamins.

Boeije, H. (2010). *Analysis in qualitative research*. Los Angeles: SAGE.

Boulton, A. (2009). Testing the limits of data-driven learning: language proficiency and training. *ReCALL (Cambridge, England), 21*(1), 37-54. doi:10.1017/S0958344009000068

Boulton, A., & Cobb, T. (2017). Corpus Use in Language Learning: A Meta-Analysis. *Language Learning, 67*(2), 348-393.

Breyer, Y. (2009). Learning and teaching with corpora: Reflections by student teachers. *Computer Assisted Language Learning, 22*(2), 153-172.

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. vi, 267 pages). Amsterdam ; Philadelphia: John Benjamins Publishing Company.

Chambers, A. (2019). Towards the corpus revolution? Bridging the research–practice gap. *Lang. Teach, 52*(4), 460-475. doi:10.1017/S0261444819000089

Cheng, W., Warren, M., & Xun-feng, X. (2003). The language learner as language researcher: putting corpus linguistics on the timetable. *System, 31*(2), 173-186. doi:http://dx.doi.org/10.1016/S0346-251X(03)00019-8

Darsih, E. (2018). LEARNER-CENTERED TEACHING: WHAT MAKES IT EFFECTIVE. *Indonesian EFL journal (Online), 4*(1), 33-42. doi:10.25134/ieflj.v4i1.796

Doyle, T. (2011). *Learner-centered teaching: putting the research on learning into practice* (1st ed ed.). Sterling, Va: Sterling, Va: Stylus Pub.

Gatto, M. (2020). Query Complexity and Query Refinement: Using Web search from a corpus perspective with digital natives. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. London: Routledge* (pp. 106-129). Oxon: Abingdon: Routledge.

Gilquin, G., & Granger, S. (2010). How can data-driven learning be used in language teaching. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (1st ed., pp. 359-370). London ; New York, NY: Routledge.

Hirata, E. (2020). The Development of a Multimodal Corpus Tool for Young EFL Learners. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. London: Routledge* (pp. 88-105). Oxon: Abingdon: Routledge.

Johns, T. (1991). "Should you be persuaded": Two samples of data-driven learning materials. *Classroom Concordancing ELR Journal, 4*, 1-16.

Kennedy, C., & Miceli, T. (2016). Cultivating effective corpus use by language learners. *Computer Assisted Language Learning, 30*(1-2), 91-114. doi:10.1080/09588221.2016.1264427

Kohn, K., Hoffstaedter, P., & Widmann, J. (2009). *BACKBONE - Pedagogic Corpora for Content & Language Integrated Learning*. Paper presented at the EUROCALL 2009 Conference, Universidad Politecnica de Valencia.

Lee, H., Warschauer, M., & Lee, J. H. (2020). Toward the Establishment of a Data-Driven Learning Model: Role of Learner Factors in Corpus-Based Second Language Vocabulary Learning. *The Modern language journal (Boulder, Colo.), 104*(2), 345-362. doi:10.1111/modl.12634

Leech, G. (1997). Teaching and language corpora: A convergence. . In A. Wichmann, S. Fligelstone, T. McEnery, & G. Knowles (Eds.), *Teaching and Language Corpora* (pp. 11-23). Harlow: Addison Wesley Longman.

Lénko-Szymánska, A. (2017). Training Teachers in Data-Driven Learning: Tackling the Challenge. *Language Learning & Technology, 21*(3), 217-214.

Lénko-Szymánska, A., & Boulton, A. (2015). Introduction: Data-driven learning in language pedagogy. In A. Lénko-Szymánska & A. Boulton (Eds.), *Multiple Affordances of Language Corpora for Data-driven Learning*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

McEnery, T., & Xiao, R. (2011). What Corpora Can Offer in Language Teaching and Learning. In R. Facchinetti (Ed.), *Corpus Linguistics 25 Years On* (pp. 364-380). Amsterdam/New York: Rodopi.

Meunier, F. (2020). A case for constructive alignment in DDL. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners*. London: Routledge.

National Research Council. (1996). *National Science Education Standards*. Washington, DC: The National Academies Press.

Papaioannou, V., Mattheoudakis, M., & Agathopoulou, E. (2020). Data-Driven Learning in a Greek Secondary Education Setting. In P. Crosthwaite (Ed.), *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. London: Routledge* (pp. 187-207). Oxon: Abingdon: Routledge.

Pérez-Paredes, P. (2019). The Pedagogical 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. London: Routledge*. Oxon: Abingdon: Routledge.

Römer, U. (2011). Corpus Research Applications in Second Language Teaching. *Annual Review of Applied Linguistics, 31*, 205-225. doi:10.1017/S0267190511000055

Shore, B. M., Birlean, C., Walker, C. L., Ritchie, K. C., LaBanca, F., & Aulls, M. W. (2009). Inquiry literacy: A proposal for a neologism. *LEARNing Landscapes, 3,* 139-155. Retrieved from http://www.learnquebec.ca/en/services /llandscapes.html

The Norwegian Directorate of Education and Training (2019). Søkertall videregående opplæring for skoleåret 2019-2020 (analyse). Retrieved from https://www.udir.no/tall-og-forskning/statistikk/statistikk-videregaende-skole/sokertall-videregaende-opplaring-for-skolearet-2019-2020/

Vyatkina, N. (2016). Data-driven learning for beginners: The case of German verb-preposition collocations. *ReCALL, 28*(2), 207-226. doi:10.1017/S0958344015000269

Walker, C. L., & Shore, B. M. (2015). Understanding Classroom Roles in Inquiry Education: Linking Role Theory and Social Constructivism to the Concept of Role Diversification. *SAGE Open, 5*(4), 215824401560758. doi:10.1177/2158244015607584

Weimer, M. (2013). *Learner-centered teaching : five key changes to practice*. In.

Wicher, O. (2020). Data-Driven Learning in the Secondary Classroom: A critical evaluation from the perspective of of foreign language didactics. In P. Crosthwaite (Ed.), *Data-*

Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners. London: Routledge.

Appendix 1: Phase 1 Interview Guide

English translation below.

Lærerintervjuer – Februar 2019

Informanter

Fire lærere i norsk videregående skole ble intervjuet. Utvalget baserte seg på lærere som hadde korpuslingvistikk som et element i lærerutdanningen sin. Lærerne underviser innenfor studiespesialiserende studier, yrkesfaglige studier, eller begge. Alle deltagende har gitt informert, skriftlig samtykke. Lærerne underviser på fire forskjellige skoler.

Forskningsspørsmål

How is corpus literacy promoted in Norwegian upper secondary schools?

Innledning (leses)

Intervjuet omhandler dine erfaringer, prosesser og holdninger til engelskfaget.

Intervjuspørsmål

0. Generelt

a. Hva gjorde at du bestemte deg for å bli engelsklærer?
b. Hva synes du er viktig i engelskundervisningen?
    - Hva ønsker du å oppnå med engelskundervisningen?

1. Læring og undervisning

a. Hvordan vil du beskrive din tilnærming til undervisning?
b. Hvilke andre aktiviteter pleier du å bruke i timene dine?
c. Hvordan tror du elever lærer best?
d. Vil du definere undervisningsmetodene dine som lærersentrert eller elevsentrert?

2. Digitalisering

a. Hvordan ser du på tilstanden til skolen du jobber på når det kommer til digitalisering?
b. Hvilken plass har teknologi og digitale verktøy i undervisningen din?
c. Hvordan arbeider du med elevene dine når de tar i bruk digitale verktøy?
d. Hvordan påvirker teknologi din undervisningspraksis? (Muligheter og utfordringer?)
e. Hvordan tror du teknologi innvirker på elevenes læring?

3. Korpuserfaring

a. Du har litt korpusbakgrunn fra utdannelsen din. Hvordan har korpus påvirket lærerpraksisen din?
b. Hvordan vil du beskrive egne korpuskunnskaper?
    - Hvordan vil du beskrive hva et korpus er?
c. På hvilke områder kan korpus være nyttig?
d. Hvordan kan elevene jobbe med korpus?
    - Direkte/indirekte bruk
e. Hvordan kan du jobbe med korpus?
f. Hvilke utfordringer ser du med bruken av korpus i undervisningen?

4. Spørreskjemaresultater

a. Majoriteten av elevene hevder å være godt vant med digitale verktøy, og at de finner det enkelt å lære seg nye digitale verktøy. Hvordan vil du beskrive de digitale ferdighetene til elevene dine?
b. Hva med ferdighetene dine?
c. Majoriteten av elever er enige i at digitale verktøy gjør det enklere for dem å lære. Hva tenker du om sammenhengen mellom digitale verktøy/teknologi og læring?
d. De fleste elevene nevner «å finne informasjon» og «å skrive/å ta notater» som måten de bruker digitale verktøy mest til. De nevner også følgende bruk av internett og søkemotorer (graf1). Hvordan arbeider du med digitale verktøy i engelskundervisningen?

e. Noen av elevene rapporterer å ha hørt om følgende korpusverktøy (graf2). Har du tatt noen av disse i bruk med elevene dine?
f. Elevene nevner nettsøk, digitale ordbøker og at de spør en medelev som strategier når de kommer over ord og uttrykk de ikke kan. Hvilke fokus har du på slike lese- og skrivestrategier i din undervisning?

English translation follows.

Teacher interviews – February 2019

Informants

Four teachers in Norwegian upper secondary school were interviewed. The sample was based on teachers who had a corpus linguistics element as part of their teacher education. The teachers taught general studies, vocational studies, or both. All participants gave their informed, written consent. The teachers worked at four different schools.

Research question

How is corpus literacy promoted in Norwegian upper secondary schools?

Introduction (read out)

The interview concerns your experiences, processes, and attitudes concerninig the English subject.

Interview questions

0. General

a. What made you choose to become an English teacher?
b. What do you think is important in English teaching and learning?
    - What do you wish to achieve with your English teaching?

1. Learning and teaching

a. How would you describe your teaching approach?
b. What sort of activities do you usually use during your lessons?
c. How do you think students learn best?
d. Would you define your teaching and learning approaches as teacher centered or learner centered?

2. Digitalization

a. What is your opinion on your school's state of digitalization?
b. What space do you give to technology and digital tools in your teaching?
c. How do you work with your students when they are using digital tools?
d. How does technology influence your teaching practice? (Possibilities and challenges?)
e. How do you think technology influences students' learning?

3. Experiences with corpora

a. You have a bit of a corpus background from your education. How have corpora influenced your teaching practice?
b. How would you describe your corpus knowledge?
    - How would you describe what a corpus is?
c. In what areas can corpora be useful?
d. How can students work with corpora?
    - Direct/indirect use.
e. How can you work with corpora?
f. What challenges do you see working with corpora in your teaching?

4. Questionnaire results (teachers are asked to comment on the student questionnaire results)

a. The majority of your students claim to be very familiar with digital tools and that they find it easy to learn new ones. How would you describe your students' digital skills?
b. How about your skills?

c. The majority of your students agree that digital tools make it easier for them to learn. What is your opinion on the connection between digital tools/technology and learning?
d. Most of the students mention “to find information” and “to write/take notes” as the areas where they use digital tools the most. They also point to the following use of the internet and search engines (show: graph 1). How do you work with digital tools in your English teaching?
e. Some of the students report to have heard about the following corpus tools (show: graph 2). Have you used any of these with them?
f. The students mention web searches, digital dictionaries and asking a peer as strategies they use when encountering words or expressions they don’t know. What focus do you have on reading and writing strategies in your lessons?

Appendix 2: Student Questionnaire

The questionnaire was completed online. The students could choose between a Norwegian and an English version. The English version follows the Norwegian one below. The following questionnaire shows the layout, items, and answer categories, but it is not the digital version the students answered.

Dette er et spørreskjema for elever som har erfaring fra Engelskfaget i norsk skole.

Ved å bruke omtrent 10 minutter på å svare på dette spørreskjemaet bidrar du til å utvikle digitale verktøy som kan være til hjelp for lærere og elever. Dette kan hjelpe lærere og elever med å utforske hvordan engelsk brukes i dagliglivet.

Deltakelse er frivillig. Informasjonen er konfidensiell, og det er kun forskeren og hans veiledere som har tilgang til den. Dataen vil bli anonymisert i alle publikasjoner. Samlet data vil bli anonymisert innen prosjektslutt (31. desember 2021). *Ansvarlig forsker: Petter Hagen Karlsen 91727680 Petter.Karlsen@inn.no*

---

1. Jeg er vant med å bruke digitale verktøy i engelskundervisningen (datamaskin, nettbrett, internett).

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

2. Jeg er vant med å bruke mobiltelefon i engelskundervisningen som en del av læringen.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

3. Jeg synes det er enkelt å lære meg nye dataprogrammer og digitale verktøy.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

4. Jeg lærer mer i undervisning hvor jeg kan bruke datamaskin, nettbrett eller andre digitale verktøy.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

5. Jeg foretrekker å jobbe med digitale hjelpemidler/verktøy, fremfor å jobbe uten.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

6. Jeg tror teknologi og digitale plattformer har gjort det enklere å søke informasjon.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

7. Jeg tror teknologi og digitale plattformer har gjort det enklere å lære seg nye ting.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

8. Jeg tror tilgangen på digitale verktøy og plattformer, som for eksempel sosiale medier (Facebook, Instagram, osv.), har gjort at jeg deltar mindre i timene.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

9. Jeg synes det er distraherende med digitale verktøy i undervisningen.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

10. Jeg skulle ønske vi brukte datamaskin, nettbrett eller mobil oftere i engelskundervisningen.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

11. Hva bruker du digitale verktøy mest til i engelskundervisningen?

Skriv svaret ditt i boksen under. Du kan skrive flere setninger.

12. Hvilke digitale verktøy og/eller nettsider bruker du på skolen?

Skriv svaret ditt i boksen under. Du kan skrive flere setninger.

13. Hvordan liker du å jobbe i timene?

NB! Du kan sette flere kryss.

- Individuelt/alene
- I par
- I grupper (3 eller flere)
- Undervisning hvor hele klassen følger med på læreren og læreren er i sentrum

14. Jeg liker undervisning som er ledet av og har fokus på læreren (tavleundervisning).

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

15. Jeg liker undervisning hvor jeg finner informasjon selv og må løse problemer på egenhånd eller i grupper.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

16. Jeg jobber mye selvstendig i engelsktimene.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

17. Jeg liker undervisning hvor jeg kan aktivt delta på diskusjoner.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

18. Jeg liker undervisning hvor jeg kan utforske ting selv.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

19. Jeg foretrekker å finne informasjon i bøker (lærebøker, ordbøker, leksikon, osv.) fremfor på nettet.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

20. Hva gjør du dersom du kommer over ord eller uttrykk på engelsk du ikke forstår?

NB! Du kan sette flere kryss.

- Jeg slår opp i papirordbøker
- Jeg søker på internett
- Jeg søker det opp i en nettordbok
- Jeg spør en medelev
- Jeg spør læreren
- Jeg hopper over ordet/uttrykket og leser videre
- Annet (fyll ut i boksen under)

21. Hva ønsker du å lære om i engelsktimene?

NB! Du kan sette flere kryss.

- Grammatikk (språkets struktur)
- Nye ord og uttrykk
- Kultur og historie fra engelsktalende land
- Litteratur og lesing
- Hvordan man kan kommunisere på engelsk
- Annet (fyll ut i boksen under)

22. Jeg synes læreren burde kunne svar på alle spørsmål jeg har om engelsk i engelsktimene.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

23. Jeg synes ikke læreren burde stille spørsmål i engelskundervisningen han/hun ikke har svaret på selv.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

24. Jeg foretrekker fasitsvar fremfor svar som ikke har en typisk fasit (diskusjonsspørsmål).

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

25. I undervisningen blir vi ofte presenter for språkeksempler faktiske engelsktalende personer.

[ ] Veldig enig[ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

26. Jeg synes det hadde vært interessant å se språkeksempler fra faktiske engelsktalende personer i undervisningen.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

27. Har du hørt om noen av nettstedene eller ressursene nevnt under?

NB! Du kan sette flere kryss.

- Sketch Engine (SKELL)
- Just-the-word.com
- AntConc
- British National Corpus (BNC)
- Corpus of Contemporary American (COCA)

- Andre nettressurser (legg inn boks)

28. Benytter du deg av noen av disse digitale ordbøkene?

NB! Du kan sette flere kryss.

- Oxford English Dictionary (www.oed.com)
- Longman Dictionary of Contemporary English
- Merriam-Webster
- Macmillian Dictionary
- Google translate/ordbok
- Ordnett
- Collins Dictionary
- Cambridge Dictionary

- Andre (fyll ut i boksen under)

29. Har du hørt om «korpus» før i sammenheng med språklæring (på engelsk: corpus / corpora)?

[ ] Ja [ ] Nei

30[7]. Hva har du hørt om korpus? (Skriv svaret ditt i boksen under)

31. Jeg trives med engelsk som fag i skolen.

[ ] Veldig enig [ ] Delvis enig [ ] Verken enig eller uenig [ ] Delvis uenig [ ] Veldig uenig

[7] Elevene fikk dette spørsmålet kun dersom de svarte 'ja' på spørsmål 29.

32. Hvordan vil du vurdere de *skriftlige* engelskkunnskapene dine basert på tilbakemeldinger du har fått fra læreren din?

[ ] Veldig gode [ ] Gode [ ] Middels [ ] Svake [ ] Veldig svake

33. Hvordan vil du vurdere de *muntlige* engelskkunnskapene dine basert på tilbakemeldinger du har fått fra læreren din?

[ ] Veldig gode [ ] Gode [ ] Middels [ ] Svake [ ] Veldig svake

34. Ser du for deg at du vil ha mer engelsk på videregående enn det obligatoriske første året?

[ ] Ja

[ ] Vet ikke

[ ] Nei

[ ] Jeg har allerede gjort det

Kjønn

[ ] Mann

[ ] Kvinne

[ ] Det ønsker jeg ikke å oppgi

[ ] Annet

Skolekode

Spør læreren din.

[ ] W1 [ ] Z1

[ ] W2 [ ] Z2

[ ] X1

[ ] X2

[ ] Y1

[ ] Y2

[ ] Y3

English version follows.

This is a questionnaire for students who have experience from English education in Norwegian schools.

By spending about 10 minutes on answering this questionnaire, you are contributing to the development of computer-based tools that can help teachers and students explore how English is used in everyday life.

Participation is voluntary. The information is confidential and only the researcher and his supervisors have access to it. The data will be anonymized in publications. Collected data will be anonymized by the end of the project (31 December 2021). *Researcher responsible: Petter Hagen Karlsen 91727680 Petter.Karlsen@inn.no*

---

1. I am used to working with digital tools in English lessons (computer, tablet, the internet).

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

2. I am used to working with my mobile phone in English lessons as part of the learning process.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

3. I find it easy to learn new computer programs and digital tools.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

4. I learn more in classes where I can use a computer, a tablet or other digital tools.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

5. I prefer working with digital aids/tools, as opposed to working without them.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

6. I think technology and digital tools have made it easier to find information.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

7. I think technology and digital tools have made it easier to learn new things.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

8. I think the access to digital tools and platforms, such as for example social media (Facebook, Instagram, etc.), has made me participate less in class.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

9. I find digital tools to be distracting in class.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

10. I wish we would use computers, tablets or mobile phones more often in English class.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

11. What do you use digital tools for the most in English class?

Click the box below to write your answer. You may write several sentences.

12. Which digital tools and/or webpages do you use at school?

Click the box below to write your answer. You may write several sentences.

13. How do you like to work in class?

NB! You can check several boxes.

- Individually/alone
- In pairs
- In groups (3 or more)
- Teaching where the entire class pays attention to the teacher and the teacher is in focus

14. I like classes that are led by and focused on the teacher ("blackboard teaching"/instruction).

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

15. I like classes where I search for information and solve problems on my own or in groups.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

16. I often work independently in English class.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

17. I like classes where I can actively participate in discussions.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

18. I like classes where I can explore things myself.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

19. I prefer to find information in books (course books, dictionaries, lexicons, etc.) as opposed to online.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

20. What do you do if you come across a word or expression in English you do not understand?

NB! You can check several boxes.

- □ I look it up in a paper dictionary
- □ I search online
- □ I look it up in an online dictionary
- □ I ask another student
- □ I ask the teacher
- □ I skip the word/expression and keep reading
- □ Other:

21. What do you want to learn about in English class?

NB! You can check several boxes.

- Grammar (the language's structure)
- New words and expressions
- Culture and history from English-speaking countries
- Literature and reading
- How to communicate in English
- Other:

22. I think the teacher should be able to answer any question I have about English during English class.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

23. I do not think the teacher should ask me questions during English class that he/she does not have the answer to himself/herself.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

24. I prefer questions with clear, definite answers (fasitsvar) as opposed to questions with more open answers (questions for discussion).

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

25. During classes, we are often presented with language examples from actual English-speaking people.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

26. I would find it interesting to see language examples in English lessons from native English speakers/speaker who use English in their daily lives.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

27. Have you heard about any of the online tools listed below?

NB! You can check several boxes.

- ☐ Sketch Engine (SKELL)
- ☐ Just-the-word.com
- ☐ AntConc
- ☐ British National Corpus (BNC)
- ☐ Corpus of Contemporary American (COCA)
- ☐ Other online and/or learning resources:

28. Do you use any of these online dictionaries?

NB! You can check several boxes.

- ☐ Oxford English Dictionary (www.oed.com)
- ☐ Longman Dictionary of Contemporary English
- ☐ Merriam-Webster
- ☐ Macmillian Dictionary
- ☐ Google translate/ordbok
- ☐ Ordnett
- ☐ Collins Dictionary
- ☐ Cambridge Dictionary
- ☐ Other dictionaries:

29. Have you heard about corpus/corpora in connection to language learning?

[ ] Yes [ ] No

30[8]. What have you heard about corpora?

[8] The students got this item if they answered 'yes' to item 29.

31. I enjoy English as a school subject.

[ ] Strongly agree [ ] Partially agree [ ] Neither agree nor disagree [ ] Partially disagree [ ] Strongly disagree

32. How would you assess your *written* English skills based on the feedback you have received from your teacher?

[ ] Very good [ ] Good [ ] Average [ ] Weak [ ] Very weak

33. How would you assess your *oral* English skills based on the feedback you have received from your teacher?

[ ] Very good [ ] Good [ ] Average [ ] Weak [ ] Very weak

34. Do you see yourself choosing more English in upper secondary school (videregående) after the obligatory first year?

[ ] Yes

[ ] I do not know

[ ] No

[ ] I have already done it

Gender:

[ ] Male

[ ] Female

[ ] I do not wish to share this information

[ ] Other:

School code:

Ask your teacher.

[ ] W1 [ ] Y1

[ ] W2 [ ] Y2

[ ] X1 [ ] Y3

[ ] X2 [ ] Z1

[ ] Z2

Appendix 3: Group Interviews

English translation below.

Intervjuguide (elever) – uke 50 & 51

Informanter

Utvalgte elevgrupper i en norsk videregående skole deltar i gruppeintervjuer. Grunnlaget for deltakelse er tidligere deltakelse i et forskningsprosjekt som omhandlet bruk av korpusbaserte læringsverktøy. Alle deltagende har gitt fritt, informert, skriftlig samtykke. Intervjuet er semi-strukturert og kan derfor gå utover guiden for å følge opp elevenes refleksjoner.

Forskningsspørsmål:

How do learners in upper secondary school experience corpus-based and corpus-aided resources?

0. Innledning:

Intervjuet omhandler dine erfaringer med prosjektperioden. Dere er selvfølgelig helt anonyme. Jeg vil gjerne stille dere noen spørsmål, og så står dere fritt til å diskutere dere imellom. Jeg håper dere kan være så ærlige og direkte som mulig. Jeg blir ikke såret om dere er negative eller positive, alt hjelper meg i dette prosjektet så si akkurat det du mener og tenker.

1. Generelt

1. Hvordan ser engelsktimene deres vanligvis ut?
2. Hva vil dere ha ut av engelskundervisningen?
3. Hvordan liker dere å jobbe i engelsktimene?
4. Hva motiverer dere til å jobbe med engelsk?
5. Hva synes dere om å bruke digitale verktøy i Engelskundervisningen?

2. Prosjektperioden

1. Hvordan har dere opplevd engelsktimene de to forrige ukene?
2. Hvilke digitale ressurser har dere jobbet med?
3. Hva var utfordrende de siste ukene?
    a. Var oppgavene utfordrende eller verktøyet?
4. Hva synes dere har vært nyttig og hva synes dere har vært unyttig i denne perioden?
5. Hvordan opplevde dere å lære å bruke dette verktøyet?
6. Hva synes dere om denne måten å jobbe på?
7. Hva kunne gjort oppgavene bedre?
8. Hva kunne gjort timene bedre?
9. Hva bidro disse verktøyene med i læringsprosessen deres?
10. Hvordan jobbet dere med oppgavene utenfor klasserommet?
11. Hvordan opplevde dere overgangen fra de konkrete oppgavene i starten, og den åpne oppgaven mot slutten?
12. Hvordan opplevde dere læreren i denne perioden?
    a. Hva gjorde læreren gjennom prosessen?
    b. Hva kan han ha gjort annerledes?

3. Spesifikke applikasjoner og temaer

1. Hva synes dere om designet på nettsiden?
2. Hvordan var det å finne frem på nettsiden?
3. Hva fikk dere ut av denne siden?
4. Hva synes dere om tekstene og videoene på nettsiden?
5. Hva synes dere om oppgavene dere fikk utdelt?
6. Hva brukte dere mest for å svare på oppgaven; *Backbone* eller andre internettressurser?
    a. Hva var grunnen til dette?
7. Tror dere at dere kommer til å bruke *Backbone* i fremtiden i eget arbeid?
8. Hvordan opplevde dere å jobbe med tekster av faktisk engelsktalende personer?
9. Benyttet dere SKELL eller Longman Dictionary of Contemporary English?
    a. Hvordan brukte dere disse ressursene?

10. Hva er et korpus?
11. Er det noe dere vil legge til eller ta opp med meg?

English translation follows.

Interview guide (students)

Informants

A selection of students in a Norwegian upper secondary school is participating in group interviews. The reason for their participation is that they partook in a research project about using a corpus-based learning tool. All participants have given their free, informed, written consent. The interviews are semi-structured and may stray from the guide to pursue students' reflections.

Research question:

How do learners in upper secondary school experience corpus-based and corpus-aided resources?

0. Introduction [read out]:

The interview is about your experiences with the project period. You will of course be completely anonymous. I want to ask you some questions, and then you are free to have a discussion among yourselves. Please be as honest and direct as possible. I will not be affected whether your responses are negative or positive; whatever you say will help the project, so speak your minds.

Interview questions:

1. General

1. What do your English lessons usually look like?
2. What do you want to get out of your English lessons?
3. How do you prefer to work during your English lessons?

4. What motivates you to work with English?
5. What do you think about working with digital tools during English lessons?

2. The project period

1. How did you experience the English lessons the previous weeks?
2. What digital resources did you use?
3. What was challenging the past two weeks?
    a. Were the tools or the tasks challenging?
4. What did you find useful and what did you find less useful during this period?
5. How did you experience using the tool?
6. What do you think about working in this manner?
7. What could have made the tasks better?
8. What could have made the lessons better?
9. How did the digital tools you used help your learning process?
10. How did you work with the tasks outside the classroom?
11. How did you experience the transition from the specific tasks at the beginning of the period, to the open writing assignments toward the end of the period?
12. How did you experience the teacher throughout this process?
    a. What did the teacher do throughout the project?
    b. What could he have done differently?

3. Specific applications & topics

1. What were your impressions of the website's design?
2. How did you experience finding your way around the website?
3. What did you get out of this website?
4. What did you think of the texts and videos on the website?
5. What did you think of the tasks you received?
6. What did you utilize most, *Backbone* or other web resources?
    a. Can you tell me why?
7. Do you think you would use *Backbone* in the future in your own work?

8. How did you experience working with texts from English-speaking people?
9. Did you make use of SKELL or the Longman Dictionary of Contemporary English?
    a. If yes, how did you use these resources?
10. What is a *corpus*?
11. Is there anything you wish to add or tell me?

Appendix 4: Case Study Lesson Plans

Class: 1a / b	**11th grade**	**English**	**Lesson duration** 90 min
LK06: • **Language learning**: evaluate different digital resources and other aids critically and independently, and use them in own language learning. • **Oral communication**: listen to and use native varieties of English from the chosen countries.			

Time	**Goal**	**Activity**	**Org.**	**Mat.**
10 min	Make introductions	Introduce the researcher and briefly explain the project.	Whole class	Nothing
55 – 60 min	Examine specific parts of language through corpora while exploring the tool	By using the first task sets (task set 1 & 2), explore the BACKBONE corpora. Look at and discuss the use and value of frequency lists in language learning. Look at and discuss the use and value of section searches in language learning. Look at and discuss the use and value of word searches in language learning. How can corpora contribute with in your learning, if at all?	Groups of 4 – 6	Tasksets 1 & 2
10 - 15 min	Summarize and discussion the lesson	Content discussion: What was covered and discovered? Meta-discussion: How did the tool work for the students?	Whole class	Nothing
10 min	Supply metadata such as English experience, age, etc.	Fill in some background information using the characteristics card template. ***No names** are to be filled in.*	Individually	Characteristics cards

Class: 1a / b	**11th grade**	**English**	**Lesson duration** 7 lessons

LK06:

- **Culture, society and literature**: Discuss and elaborate on culture and social conditions in two self-chosen English-speaking countries.
- **Oral communication**: listen to and use native varieties of English from the chosen countries.
- **Written communication**: use conventions for English language construction in order to communicate effectively in writing.
- **Written communication**: In reading and gathering material, the pupil needs to evaluate the content from sources in an independent, critical and verifiable way.
- **Language learning**: evaluate different digital resources and other aids critically and independently, and use them in own language learning.

Time	**Goal**	**Activity**	**Org.**	**Mat.**
10 min	Introduce the project	The students will work in groups on **<u>one</u>** of two tasks. The tasks' design is inspired by previous exam tasks. Go through the project structure: a) 2 lessons will be spent on examining the corpus and discussing their findings. b) 5 lessons will be spent creating a text based on the handout. c) 1 lesson will be spent discussing the tasks, their findings, and the tools used. Hand out the project tasks.	Whole class	Project handout
60 min	Explore the project and find information	Start exploring the corpus and available resources.	Individually or in smaller groups	Project handout
20 min	Discuss your choices and finding in groups	Go into groups based on which task was chosen (same task = same group). Discuss preliminary findings and comment on each other's work.	Groups of 5 - 6	Project handout
5 lessons	create a text using the corpus and other resources	Follow the task instructions (project handout) to create a text.	Groups of 5 - 6	Project handout
1 lesson	Discuss the period and its value	Summary: Discuss with the whole class the findings of the projects. Evaluate the tools used for the project.	Whole class	Everything

Appendix 5: *Backbone* Task-sets

User manual for BACKBONE (BB)

Access the website here: http://webapps.ael.uni-tuebingen.de/backbone-search/faces/initialize.jsp

1. Choose a corpus

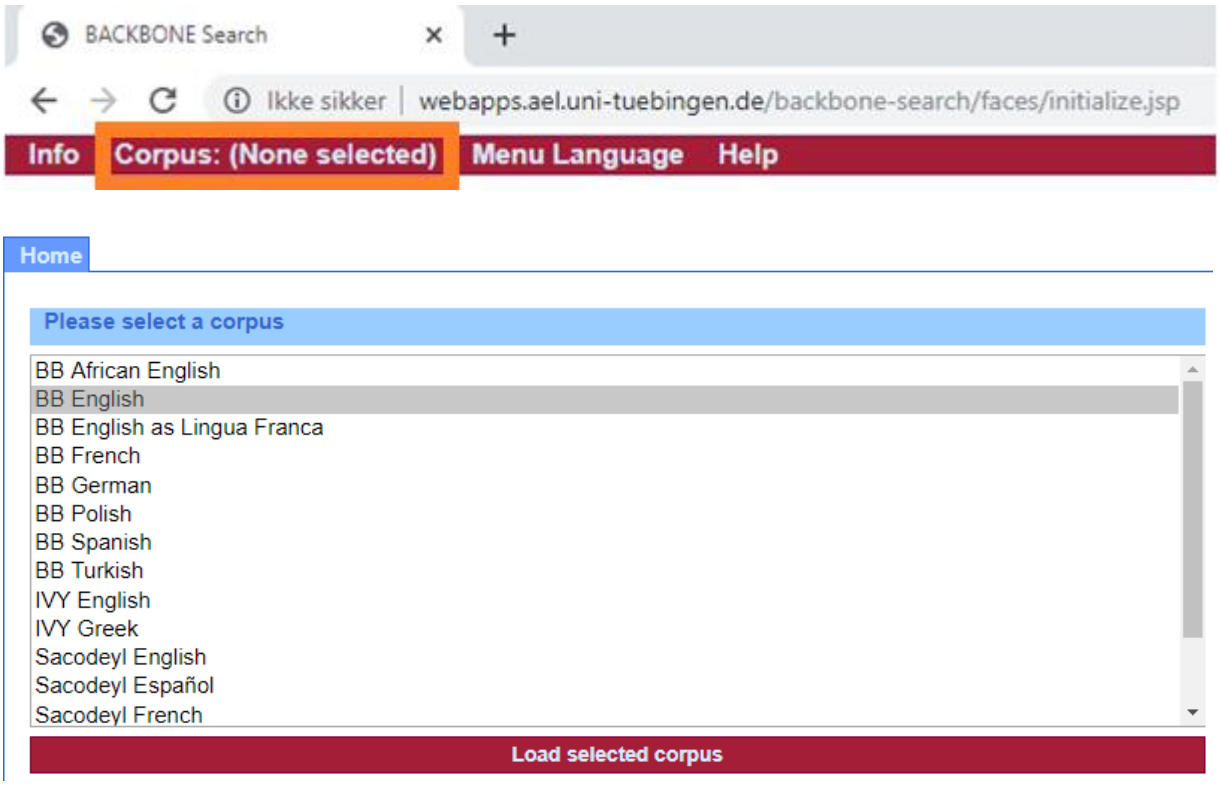


Table 1.

- Here you can see the different *corpora* (plural of *corpus*). A corpus is a collection of texts or, in this case, videos and texts.
- By clicking one of the lines and then *Load selected corpus* you access one of these collections.

- You can change the corpus you are working with at any time by clicking the button indicated by the orange box. This will give you access to different interviews and texts.

2. Switching corpora

a) Choose *IVY English* from the list.
b) The *browse* button lets you see each interview (at the top of the screen).
c) What sort of videos do you see? Discuss. Look at some of the videos by pressing *Play video* to the right of the screen.
d) Press "transcriptions" to the right of the interviews. What do you see?
e) Try switching to different corpora with the **button marked in orange** in Table 1.
f) Look at and discuss what types of texts are in a couple of different corpora.

3. The words of the corpus

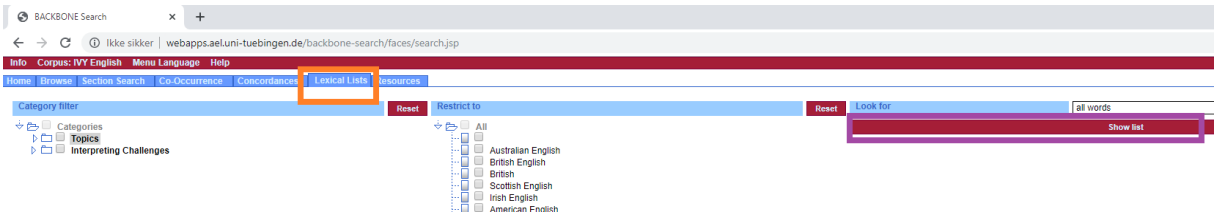


Table 2.

a) Go to *Lexical Lists* indicated in **orange** in Table 2.
b) Press *Show list* indicated in **purple**.
c) You now see all the words in the corpus and how many times they appear in all the interviews. This is called ***frequency***. Why do you think the list looks like it does?
d) Discuss what words you think are used when talking about the two topics *culture* and *education* before searching the corpus.
e) Compare the two lists by pressing *topic*, then mark the topic you want to look at and press *show list*. Were the lists as expected? Any differences or similarities?
f) Compare some other lists and discuss your results.

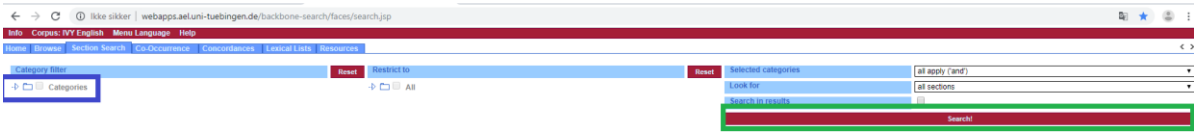
4. Looking into idioms

a) Do you know what an *idiom* is? Do you know any in English or your native language? Discuss it in groups before going to the computer.

b) Go to Longman Dictionary of Contemporary English and search "idiom". What are the definition and some examples?

c) Go to the corpus and follow the guide below in this document (see Table 3).

- Load either the IVY corpus or the SACODEYL corpus.
- Press "Section Search" (**blue box**): this searches ALL the interviews in the corpus at once.
- Press Categories → Interpreting Challenges → Lexis → Idiomatic Phrases → Search (**green box**).
- Different sections of each interview pop up below! Press *Interpreting challenges – Idiomatic phrases* under *Annotations* (**purple box**) on each interview, then *Highlight selected annotations*. The idioms will now be **red** in the text!



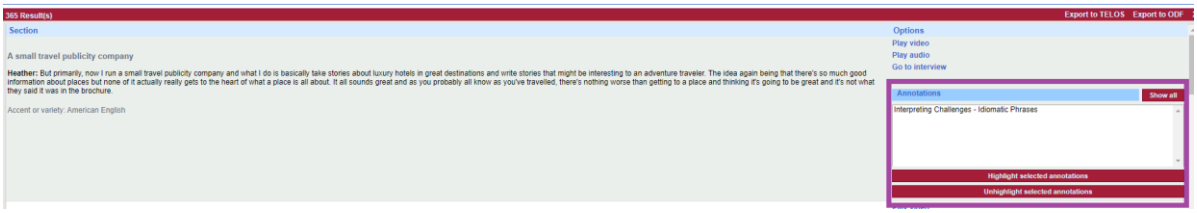


Table 3

d) Find 7 idioms from different sections (you can scroll down to see more sections).

e) Discuss their meaning in your group (before searching the web). Are they the same in your native language if you translate them?

f) If you don't know, you can search the web to find out.

g) Present your findings to another group.

h) What other types of topics and language can you search for under "categories"? Explore!

Tasks for BACKBONE (BB)

Access the website here: http://webapps.ael.uni-tuebingen.de/backbone-search/faces/initialize.jsp

1. Exploring an interview

a) Go to *Browse* and choose an interview (see Table 1).

b) Listen to the interview and discuss what the person is talking about.

c) What varieties of English are there in the IVY corpus? Listen to a few. Have you heard any of them before? Choose one and discuss your choice.

d) Press *Show transcription* to the right of an interview (**green box**) to get the interview in text form.

e) See the different categories in the interview under *Annotation* (**orange box**).

f) Explore a couple of the categories by pressing them, and then *Highlight annotation*.

g) How can this tool help you when writing a text?

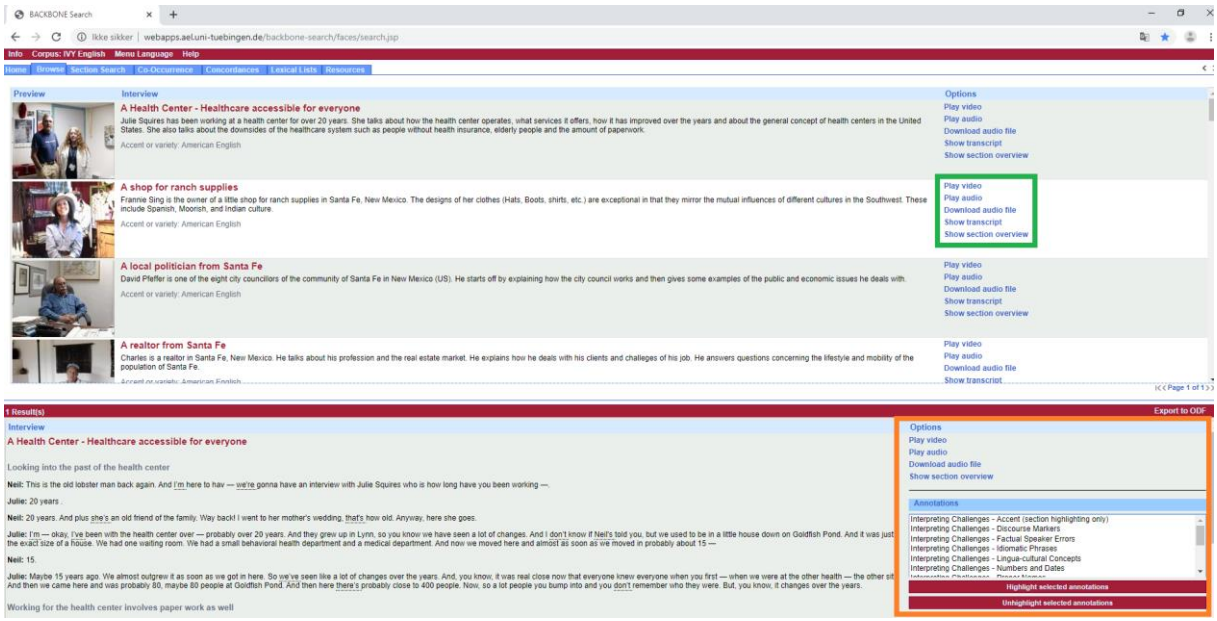


Table 1.

2. Finding words and their contexts

a) You can also search for how different words are used in the corpus.

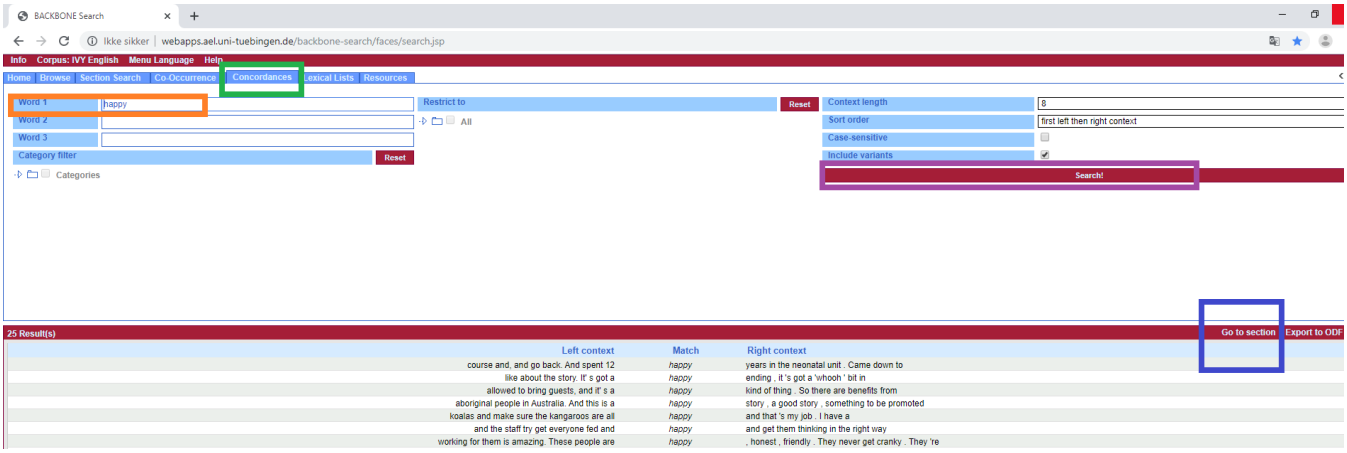
b) Go to *Concordances* (**blue box**) and **search** for the word "happy" (**orange box**).

c) The word appears below from all the interviews in the corpus. On each side of it, you can see the context (see Table 2). How is it used?

d) To see who talks about "happy", select one of the lines and press *Go to section* (**blue box**) to go to the interview.

e) What can we do with this information?

f) Try some different searches! Explore the tool! What else can it do?



Want to see more examples of how words are used?

Go to SKELL: https://skell.sketchengine.co.uk/run.cgi/skell

Search for any word and see it in context, its synonyms, or what words it tends to hang out with.

Choose <u>one</u> of these tasks for your project:

Task 1. Education and Gender

Create a text discussing the issue of gender in schools in English-speaking countries. Use the SACODEYL corpus interview of Helen as a starting point for your discussion.

- Listen to the interview of Helen in the SACODEYL corpus.
- Use the corpus to explore how Helen talks about the topic *Education and gender*.
- Can you find other references to gender or educational issues?
- Discuss her viewpoints and use other sources from the internet; remember to reference your work.
- What are your opinions on the issue?

Tips!

- **Lexical lists:** You can for example use the lexical list for *<u>school</u>* to see typical words used to talk about the topic, or any other topic.
- **Section search:** You can search for different ways of writing and word use by searching the different categories (see the *help section* on the next page).
- **Concordances:** You can search for how a specific word is used in all the different texts.

Task 2. Environmental Issues

Create a text discussing Environmental issues in English-speaking countries. Use the IVY corpus interviews as a starting point for your discussion.

- Use the *Section search* in the IVY corpus to look for the topic *Ecological Issues.*
- Use the corpus to explore how different people talk about different ecological issues.
- Choose a couple of interviews and watch them.
- Discuss their viewpoints and use other sources from the internet; remember to reference your work.
- What are your opinions on the issue?

Tips!

- **Lexical lists:** You can for example use the lexical list for *<u>environment</u>* to see typical words used to talk about the topic.
- **Section search:** You can search for different ways of writing and word use by searching the different categories (see the *help section* on the next page).
- **Concordances:** You can search for how a specific word is used in all the different texts.

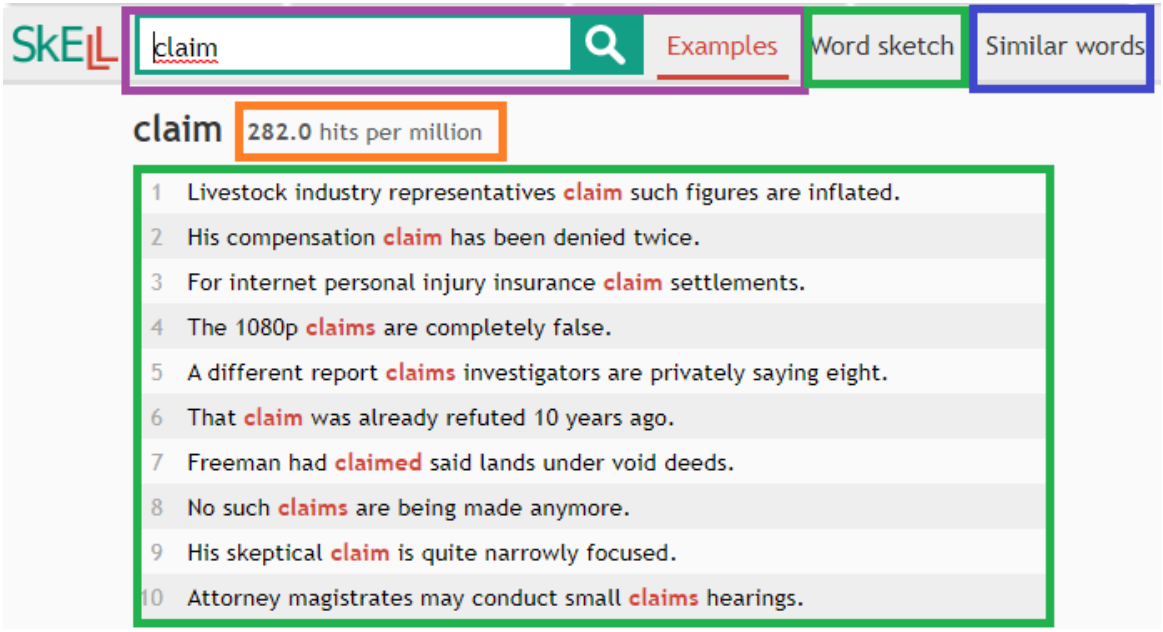
Help Section

How to reference corpora

In the text (examples)	In your reference list
James tells us that… (SACODEYL)	SACODEYL. (2009-2010). System aided compilation and open distribution of European youth language. Retrieved from: http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp
James tells us that… (IVY)	IVY. (2009-2010). IVY English corpus. Retrieved from: http://webapps.ael.uni-tuebingen.de/backbone-search/faces/search.jsp

How to use SKELL to find ways of writing

1. Go to: https://skell.sketchengine.co.uk/run.cgi/skell
2. This website lets you see words in context, how they can work in a sentence, and their synonyms. SKELL uses the British National Corpus (BNC), which is a huge corpus of British English.
3. For instance, when quoting someone's statement about a topic, you can write: *James claims…* or *Julia suggests…* If you go to SKELL, you can search for these verbs and find out how they are used and if they are appropriate for your use (**purple box**).
4. You can see what purpose they have in a sentence or if they have different purposes (**green box**);
5. how much they are used per million words, which means that if you take a million words from the corpus, "claim" is 282 of these words (**orange box**);
6. or you can see other words that mean almost the same (**blue box**).
7. Since "claim" is both a verb and a noun, you can see both below. Go to "word sketch" (**green box**) to see only verb or only noun.



Appendix 6: NSD Evaluation

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

Corpora in the EFL Classroom

Referansenummer

140840

Registrert

21.08.2018 av Petter Hagen Karlsen - petter.karlsen@inn.no

Behandlingsansvarlig institusjon

Høgskolen i Innlandet / Fakultet for lærerutdanning og pedagogikk / Institutt for humanistiske fag

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Petter Hagen Karlsen, Petter.Karlsen@inn.no, tlf: 91727680

Type prosjekt

Forskerprosjekt

Prosjektperiode

01.08.2018 - 31.12.2021

Status

30.04.2021 - Vurdert

30.04.2021 - Vurdert

Vi viser til endring registrert 11.02.2020 og melding 15.04.2021. Vi kan ikke se at det er gjort noen oppdateringer i meldeskjemaet eller vedlegg som har innvirkning på NSD sin vurdering av hvordan personopplysninger behandles i prosjektet.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til videre med prosjektet!

23.10.2019 - Vurdert

NSD har vurdert endringen registrert 16.10.2019.

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 23.10.2019. Behandlingen kan fortsette.

Endringen gjelder at det skal gjennomføres et gruppeintervju med 4-6 elever fra klasse 1 ved videregående 2, samt et gruppeintervju med 4-6 elever fra klasse 2 ved videregående skole 2. De oppdaterte informasjonsskrivene er godt utformet.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Kontaktperson hos NSD: Lise A. Haveraaen

Tlf. Personverntjenester: 55 58 21 17 (tast 1)

05.11.2018 - Vurdert

Det er vår vurdering at behandlingen av personopplysninger i prosjektet vil være i samsvar med personvernlovgivningen så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet med vedlegg den 05.11.2018, samt i meldingsdialogen mellom innmelder og NSD. Behandlingen kan starte.

MELD ENDRINGER
Dersom behandlingen av personopplysninger endrer seg, kan det være nødvendig å melde dette til NSD ved å oppdatere meldeskjemaet. På våre nettsider informerer vi om hvilke endringer som må meldes. Vent på svar før endringer gjennomføres.

TYPE OPPLYSNINGER OG VARIGHET
Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til 31.12.2021.

LOVLIG GRUNNLAG
Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake. Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER
NSD vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om ogsamtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte ogberettigede formål, og ikke behandles til nye, uforenlige formål
- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendigefor formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylleformålet

DE REGISTRERTES RETTIGHETER
Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: åpenhet (art. 12), informasjon (art. 13), innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), underretning (art. 19), dataportabilitet (art. 20).

NSD vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER
NSD legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

Checkbox er databehandler i prosjektet. NSD legger til grunn at behandlingen oppfyller kravene til bruk av databehandler, jf. art 28 og 29.

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

OPPFØLGING AV PROSJEKTET
NSD vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Kontaktperson hos NSD: Lise Aasen Haveraaen
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Appendix 7: Information and Consent Forms

For teachers, research phase 1

Vil du delta i forskningsprosjektet

"Corpora in the EFL Classroom"?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å utforske elever og læreres opplevelse og oppfatning av digitale språkverktøy og undervisningsmetoder i engelskundervisningen. I dette skrivet gir jeg deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Forskningsprosjektet er en del av et doktorgradsprosjekt som strekker seg over en treårsperiode hvor formålene er å utforske elevers tidligere erfaringer med digitale verktøy og undervisningsmetoder, herunder noen spesifikke verktøy og metoder. Noen spesifikke digitale verktøy og undervisningsmetoder vil deretter bli prøvd ut i engelskundervisningen for å utforske elevers og læreres opplevelse av og interaksjon med disse verktøyene og metodene.

Hvem er ansvarlig for forskningsprosjektet?

Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?

Du er valgt ut fordi du underviser på et klassetrinn som er av spesiell interesse for forskningen som er nevnt ovenfor.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det at du deltar på et intervju om dine og klassens arbeidsmetoder og bruk av digitale verktøy i tidligere undervisning, samt holdninger i forbindelse med disse. Jeg tar lydopptak og notater fra intervjuet.

Jeg vil også be elvene dine fylle ut et spørreskjema. Det vil være ganske kort og vil ikke ta lang tid (omtrent ti minutter). Spørreskjemaet inneholder spørsmål om tidligere erfaringer fra engelskundervisningen de har hatt, som for eksempel bruk av digitale verktøy. Dine svar fra spørreskjemaet blir registrert elektronisk.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Det vil ikke påvirke ditt forhold til skolen/lærer.

Ditt personvern – hvordan jeg oppbevarer og bruker dine opplysninger

Jeg vil bare bruke opplysningene om deg til formålene jeg har fortalt om i dette skrivet. Jeg behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Kun forskeren selv (undertegnede) og eventuelt veiledere vil ha tilgang til opplysningene dine.
- For å forhindre at andre får tilgang så vil personopplysninger om deg (f. eks. navn og alder) bli erstattet med en kode som bare forskeren har tilgang til. Dokumentet med denne koden og opplysningene om deg vil oppbevares separat og innelåst.
- I tillegg oppbevares alle papirer og datamaskiner innelåst. Ingen av opplysningene om deg vil bli lagret på private datamaskiner eller minnepinner.

Det vil ikke bli mulig å gjenkjenne deg i publikasjonene denne informasjonen blir brukt til. Den eneste personinformasjonen som vil komme frem i publikasjonene er alder/klassetrinn, kjønn, og type skole (ungdomsskole, videregående, etc.).

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes 31.12.21. Når prosjektet blir avsluttet slettes kodene som knytter navnet ditt til dataen, og blir derfor helt anonymisert.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir meg rett til å behandle personopplysninger om deg?
Jeg behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk ved Petter Hagen Karlsen via epost: petter.karlsen@inn.no eller telefon: 91 72 76 80.
- NSD – Norsk senter for forskningsdata AS, på epost (personvernombudet@nsd.no) eller telefon: 55 58 21 17.
- Høgskolen i Innlandets lokale kontaktperson for personvern i forskning: Anne Sofie Lofthus, anne.lofthus@inn.no, telefon: 61288277

Med vennlig hilsen

Petter Hagen Karlsen
Prosjektansvarlig
Stipendiat ved Høgskolen i Innlandet

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet Corpora in the EFL Classroom, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i intervju med lydopptak
- ☐ at elever kan gi opplysninger om meg til prosjektet

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. 31.12.21.

(Signert av prosjektdeltaker, dato)

For teachers, research phase 1 & 2

Vil du delta i forskningsprosjektet

"Corpora in the EFL Classroom"?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å utforske elever og læreres opplevelse og oppfatning av digitale språkverktøy og undervisningsmetoder i engelskundervisningen. I dette skrivet gir jeg deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Forskningsprosjektet er en del av et doktorgradsprosjekt som strekker seg over en treårsperiode hvor formålene er å utforske elevers tidligere erfaringer med digitale verktøy og undervisningsmetoder, herunder noen spesifikke verktøy og metoder. Noen spesifikke digitale verktøy og undervisningsmetoder vil deretter bli prøvd ut i engelskundervisningen for å utforske elevers og læreres opplevelse av og interaksjon med disse verktøyene og metodene.

Hvem er ansvarlig for forskningsprosjektet?

Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?

Du er valgt ut fordi du underviser på et klassetrinn som er av spesiell interesse for forskningen som er nevnt ovenfor.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det at elevene dine fyller ut to spørreskjemaer, at du blir observert/filmet i undervisningen, og at du deltar på to intervjuer. Det første intervjuet vil være ganske kort og vil ikke ta lang tid. Intervjuet vil dreie seg om dine erfaringer med engelskundervisningen og digitale verktøy.

Neste steg innebærer at disse digitale ressursene og metodene blir prøvd ut i din klasse. Denne utprøving vil strekke seg over flere uker og undervisningstimer. Hensikten med dette er å se hvordan ressursene og metodene fungerer i det enkelte klasserom. Det kan være at forsker utvikler klasseromsaktivitetene i samarbeid med deg, for å gjøre det mest mulig hensiktsmessig for både deg og elevene dine. Undervisningen vil bli observert av forskeren og/eller filmet. Kun jeg og veiledere vil ha tilgang på innspillingene. Observasjonen vil dreie seg om hvordan klassen og læreren jobber med verktøyet i praksis, og opplysninger sånn som uttalelser om prosessen eller verktøyet, og måter du tilnærmer deg verktøyet på i samhandling med elevene dine underveis, vil bli registrert.

Siste steg i prosjektet består av et nytt intervju som lar deg dele dine meninger og erfaringer rundt bruken av de nye verktøyene og metodene du har opplevd. Dine svar fra intervjuet blir spilt inn via lydopptak og det vil bli tatt notater underveis.

Jeg vil også be dine elever gi noen opplysninger om deg i et intervju i sluttfasen av prosjektet. Det vil være opplysninger om undervisning du har hatt tidligere. Jeg tar lydopptak og notater fra intervjuet.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Det vil ikke påvirke ditt forhold til skolen/lærer.

Ditt personvern – hvordan jeg oppbevarer og bruker dine opplysninger

Jeg vil bare bruke opplysningene om deg til formålene jeg har fortalt om i dette skrivet. Jeg behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Kun forskeren selv (undertegnede) og eventuelt veiledere vil ha tilgang til opplysningene dine.
- For å forhindre at andre får tilgang så vil personopplysninger om deg (f. eks. navn og alder) bli erstattet med en kode som bare forskeren har tilgang til. Dokumentet med denne koden og opplysningene om deg vil oppbevares separat og innelåst.
- I tillegg oppbevares alle papirer og datamaskiner innelåst. Ingen av opplysningene om deg vil bli lagret på private datamaskiner eller minnepinner.

Det vil ikke bli mulig å gjenkjenne deg i publikasjonene denne informasjonen blir brukt til. Den eneste personinformasjonen som vil komme frem i publikasjonene er alder/klassetrinn, kjønn, og type skole (ungdomsskole, videregående, etc.).

Hva skjer med opplysningene dine når vi avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes 31.12.21. Når prosjektet blir avsluttet slettes kodene som knytter navnet ditt til dataen, og blir derfor helt anonymisert.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir meg rett til å behandle personopplysninger om deg?
Jeg behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk ved Petter Hagen Karlsen via epost: petter.karlsen@inn.no eller telefon: 91 72 76 80.
- NSD – Norsk senter for forskningsdata AS, på epost (personvernombudet@nsd.no) eller telefon: 55 58 21 17.
- Høgskolen i Innlandets lokale kontaktperson for personvern i forskning: Anne Sofie Lofthus, anne.lofthus@inn.no, telefon: 61288277

Med vennlig hilsen

Petter Hagen Karlsen
Prosjektansvarlig
Stipendiat ved Høgskolen i Innlandet

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet Corpora in the EFL Classroom, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i intervju med lydopptak
- ☐ å delta i/gjennomføre undervisning hvor jeg blir observert/filmet
- ☐ at elever kan gi opplysninger om meg til prosjektet

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. 31.12.21.

(Signert av prosjektdeltaker, dato)

For students, research phase 1

Vil du delta i forskningsprosjektet

"Corpora in the EFL Classroom"?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å utforske elever og læreres opplevelse og oppfatning av digitale språkverktøy og undervisningsmetoder i engelskundervisningen. I dette skrivet gir jeg deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Forskningsprosjektet er en del av et doktorgradsprosjekt som strekker seg over en treårsperiode hvor jeg skal prøve ut noen digitale språkverktøy og læringsmetoder i engelskundervisningen. Formålet for denne delen av studien er å utforske elevers tidligere erfaringer og preferanser når det kommer til digitale verktøy og undervisningsmetoder, og deres kjennskap til noen spesifikke digitale språk- og læringsverktøy.

Opplysningene om deg vil bli brukt i de andre delene av doktorgradsprosjektet. Du vil være anonym i publikasjonene fra denne forskningen, det vil altså ikke være mulig å identifisere deg som enkeltperson i artiklene som skrives i dette prosjektet.

Hvem er ansvarlig for forskningsprosjektet?

Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?

Det er tre lærere ved forskjellige skoler og fem til seks klasser som er bedt om å delta på prosjektet. Du er valgt ut fordi jeg har kommet med en forespørsel til din lærer, som takket ja til å delta i prosjektet, og fordi du er på et klassetrinn som passer prosjektet. Det er selvsagt frivillig for deg å delta.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det at du fyller ut et spørreskjema. Det vil være ganske kort og vil ikke ta deg lang tid (omtrent ti minutter). Spørreskjemaet inneholder spørsmål om tidligere erfaringer fra engelskundervisningen du har hatt, som for eksempel bruk av digitale verktøy. Dine svar fra spørreskjemaet blir registrert elektronisk.

Jeg vil også be din lærer om å gi noen opplysninger om klassens arbeidsmetoder og bruk av digitale verktøy i tidligere undervisning i et intervju. Det vil ikke være snakk om opplysninger om deg som enkeltelev, men om klassen generelt. Jeg tar lydopptak og notater fra intervjuet.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Det vil ikke påvirke ditt forhold til skolen/lærer.

Ditt personvern – hvordan jeg oppbevarer og bruker dine opplysninger

Jeg vil bare bruke opplysningene om deg til formålene jeg har fortalt om i dette skrivet. Jeg behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Kun forskeren selv (undertegnede) og eventuelt veiledere vil ha tilgang til opplysningene dine.
- For å forhindre at andre får tilgang så vil personopplysninger om deg (f. eks. navn og alder) bli erstattet med en kode som bare forskeren har tilgang til. Dokumentet med denne koden og opplysningene om deg vil oppbevares separat og innelåst.
- I tillegg oppbevares alle papirer og datamaskiner innelåst. Ingen av opplysningene om deg vil bli lagret på private datamaskiner eller minnepinner.
- Data om deg som blir samlet inn via spørreskjema vil bli samlet inn og lagret i Høgskolen i Innlandets server gjennom spørreskjematjenesten Checkbox. Serveren benytter ikke en skytjeneste og dette vil medføre sikker behandling av dataen.

Det vil ikke bli mulig å gjenkjenne deg i publikasjonene denne informasjonen blir brukt til. Den eneste personinformasjonen som vil komme frem i publikasjonene er alder/klassetrinn, kjønn, og type skole (ungdomsskole, videregående, etc.).

Hva skjer med opplysningene dine når jeg avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes 31.12.21. Når prosjektet blir avsluttet slettes kodene som knytter navnet ditt til dataen, og blir derfor helt anonymisert.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir meg rett til å behandle personopplysninger om deg?
Jeg behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?
Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk ved Petter Hagen Karlsen via epost: petter.karlsen@inn.no eller telefon: 91 72 76 80.
- NSD – Norsk senter for forskningsdata AS, på epost (personvernombudet@nsd.no) eller telefon: 55 58 21 17.
- Høgskolen i Innlandets lokale kontaktperson for personvern i forskning: Anne Sofie Lofthus, anne.lofthus@inn.no, telefon: 61288277

Med vennlig hilsen

Petter Hagen Karlsen
Prosjektansvarlig
Stipendiat ved Høgskolen i Innlandet

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet Corpora in the EFL Classroom, og har fått anledning til å stille spørsmål. Jeg samtykker til:

☐ å delta i en spørreundersøkelse (svare på et spørreskjema)

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. 31.12.21.

(Signert av prosjektdeltaker, dato)

For students, research phase 2

Vil du delta i forskningsprosjektet

"Corpora in the EFL Classroom"?

Dette er et spørsmål til deg om å delta i et forskningsprosjekt hvor formålet er å utforske elever og læreres opplevelse og oppfatning av digitale språkverktøy og undervisningsmetoder i engelskundervisningen. I dette skrivet gir jeg deg informasjon om målene for prosjektet og hva deltakelse vil innebære for deg.

Formål

Forskningsprosjektet er en del av et doktorgradsprosjekt som strekker seg over en treårsperiode hvor formålene er å utforske elevers tidligere erfaringer med digitale verktøy og undervisningsmetoder, herunder noen spesifikke verktøy og metoder. Noen spesifikke digitale verktøy og undervisningsmetoder vil deretter bli prøvd ut i engelskundervisningen for å utforske elevers og læreres opplevelse av og interaksjon med disse verktøyene og metodene.

Hvem er ansvarlig for forskningsprosjektet?

Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk er ansvarlig for prosjektet.

Hvorfor får du spørsmål om å delta?

Det er tre lærere ved forskjellige skoler og fem til seks klasser som er bedt om å delta på prosjektet. Du er valgt ut fordi jeg har kommet med en forespørsel til din lærer, som takket ja til å delta i prosjektet, og fordi du er på et klassetrinn som passer prosjektet. Det er selvsagt frivillig for deg å delta.

Hva innebærer det for deg å delta?

Hvis du velger å delta i prosjektet, innebærer det at du fyller ut to spørreskjemaer, blir observert/filmet i undervisningen, og potensielt blir spurt om å delta på et intervju. Det første spørreskjemaet vil være ganske kort (omtrent en side) og vil ikke ta deg lang tid. Spørreskjemaet inneholder spørsmål om tidligere erfaringer fra engelskundervisningen du har hatt, som for eksempel bruk av digitale verktøy. Dine svar fra spørreskjemaet blir registrert elektronisk.

Neste steg innebærer at disse digitale ressursene og metodene blir prøvd ut i din klasse. Denne utprøving vil strekke seg over flere uker og undervisningstimer. Hensikten med dette er å se hvordan ressursene og metodene fungerer i det enkelte klasserom. Undervisningen vil bli observert av forskeren eller filmet. Kun jeg og veiledere vil ha tilgang på innspillingene. Observasjonen vil dreie

seg om hvordan klassen og læreren jobber med verktøyet i praksis, og opplysninger sånn som uttalelser om prosessen eller verktøyet, måter du bruker verktøyet til å søke opp informasjon og arbeide med oppgaver på, og tekster/svar du kommer frem til underveis vil bli registrert. Dersom du ikke ønsker å delta vil det bli mulig for deg å få et alternativt opplegg i et annet rom, slik at du både slipper å delta og å bli filmet uten å miste undervisning.

Siste steg i prosjektet består av et nytt spørreskjema som lar deg dele dine meninger og erfaringer rundt bruken av de nye verktøyene og metodene du har opplevd. Dine svar fra spørreskjemaet blir registrert elektronisk. Noen vil også bli spurt om å delta i et intervju. Dette er selvsagt helt frivillig. I intervjuet vil jeg spørre deg om din opplevelse av og synspunkter på arbeidet med de digitale ressursene og metodene. Jeg tar lydopptak og notater fra intervjuet som kun jeg og veilederne mine har tilgang til. En egen samtykkeerklæring for intervju kommer dersom det blir aktuelt å spørre deg.

Jeg vil også be din lærer om å gi noen opplysninger om klassens arbeidsmetoder og bruk av digitale verktøy i tidligere undervisning i et intervju. Det vil ikke være snakk om opplysninger om deg som enkeltelev, men om klassen generelt. Jeg tar lydopptak og notater fra intervjuet.

Det er frivillig å delta
Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykke tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Det vil ikke påvirke ditt forhold til skolen/lærer.

Ditt personvern – hvordan jeg oppbevarer og bruker dine opplysninger

Jeg vil bare bruke opplysningene om deg til formålene jeg har fortalt om i dette skrivet. Jeg behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

- Kun forskeren selv (undertegnede) og eventuelt veiledere vil ha tilgang til opplysningene dine.
- For å forhindre at andre får tilgang så vil personopplysninger om deg (f. eks. navn og alder) bli erstattet med en kode som bare forskeren har tilgang til. Dokumentet med denne koden og opplysningene om deg vil oppbevares separat og innelåst.
- I tillegg oppbevares alle papirer og datamaskiner innelåst. Ingen av opplysningene om deg vil bli lagret på private datamaskiner eller minnepinner.
- Data om deg som blir samlet inn via spørreskjema vil bli samlet inn og lagret i Høgskolen i Innlandets server gjennom spørreskjematjenesten Checkbox. Serveren benytter ikke en skytjeneste og dette vil medføre sikker behandling av dataen.

Det vil ikke bli mulig å gjenkjenne deg i publikasjonene denne informasjonen blir brukt til. Den eneste personinformasjonen som vil komme frem i publikasjonene er alder/klassetrinn, kjønn, og type skole (ungdomsskole, videregående, etc.).

Hva skjer med opplysningene dine når jeg avslutter forskningsprosjektet?

Prosjektet skal etter planen avsluttes 31.12.21. Når prosjektet blir avsluttet slettes kodene som knytter navnet ditt til dataen, og blir derfor helt anonymisert.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir meg rett til å behandle personopplysninger om deg?

Jeg behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk har NSD – Norsk senter for forskningsdata AS vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Hvor kan jeg finne ut mer?

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med:

- Høgskolen i Innlandet, fakultet for lærerutdanning og pedagogikk ved Petter Hagen Karlsen via epost: petter.karlsen@inn.no eller telefon: 91 72 76 80.
- NSD – Norsk senter for forskningsdata AS, på epost (personvernombudet@nsd.no) eller telefon: 55 58 21 17.
- Høgskolen i Innlandets lokale kontaktperson for personvern i forskning: Anne Sofie Lofthus, anne.lofthus@inn.no, telefon: 61288277

Med vennlig hilsen

Petter Hagen Karlsen
Prosjektansvarlig
Stipendiat ved Høgskolen i Innlandet

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet Corpora in the EFL Classroom, og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ å delta i spørreundersøkelse (svare på et spørreskjema)
- ☐ å delta i undervisning hvor jeg blir observert/filmet
- ☐ å delta på et gruppeintervju

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. 31.12.21.

(Signert av prosjektdeltaker, dato)

Inland Norway
University of
Applied Sciences

This doctoral dissertation presents the use of corpus-based approaches to English language learning in upper secondary school in Norway. The research was conducted in two distinct phases. The first phase investigated the pedagogic corpus work of four corpus-trained, in-service teachers and their students' corpus experience alongside factors that might have influenced this work. Data were collected through a questionnaire to the students and teacher interviews. The second phase featured a teacher-researcher collaboration with one teacher from the first research phase and two of his upper secondary English classes where we designed and implemented a corpus-based approach in the classroom. Data were collected through a case study design with classroom observations and subsequent student group interviews.

The findings show that the teachers of the first phase had avoided corpus-based approaches in their practice, and few of their students knew anything about corpora. Several learning opportunities were observed in the second phase including instances of metatalk to describe corpus data, and peer scaffolding to learn the tool. However, the students' impressions were negatively skewed. Several obstacles to successful corpus use in the classroom from both phases were found and fell within two broad categories. These categories concern the novelty of the approach and the training and mediation required to overcome it, and the relevance of the approach to the teachers, students, and the curriculum. Inquiry-based education was applied as a theoretical framework that has considerable overlap with the concepts of corpus-based approaches in the classroom but includes a more pronounced social dimension that foster teacher and peer mediation, collaborative learning, and knowledge sharing. It is therefore argued that viewing corpus-based approaches to the classroom as a mode of inquiry-based education can help alleviate the aforementioned issues related to novelty and relevance.