Korpusbasert undervisning i fremmedspråkene: La elevens nysgjerrighet sette dagsorden.

*Sammendrag*

Korpus er en type digitale språkressurser (maskinlesbare tekstsamlinger) som er mye brukt i moderne lingvistikk som empirisk støtte for studiet av ulike språklige fenomener. Korpus kan med hell brukes i fremmedspråkundervisningen, men er frem til nå ikke i tilstrekkelig grad blitt tatt i bruk i norsk skole og høyere utdanning. I denne artikkelen gir vi en gjennomgang av eksisterende forskningslitteratur på feltet, med hovedvekt på bidrag fra det angloamerikanske språkområdet. Deretter går vi gjennom to korpus som er tilgjengelige for norske brukere,*Norwegian-Spanish Parallel Corpus* og *Oslo Multilingual Corpus*. Vi viser hvordan disse korpusene kan brukes i undervisningen av ulike språklige fenomener, fra vokabular til sosiolingvistiske fenomener.

På grunnlag av eksisterende forskning på pedagogiske applikasjoner av korpus samt våre egne forslag til undervisningsopplegg er det grunn til å tro at bruk av korpus vil kunne være et verdifullt bidrag til fremmedspråkundervisningen i norske klasserom. Likevel finnes visse utfordringer som hindrer at korpus tas i bruk. Både mangel på tid, gruppestørrelser og teknologiske hindre kan stå i veien. Bruk av korpus i fremtidens klasserom er også nært knyttet til ulike trender, både innenfor teknologi og forskning. For det første blir datateknologi stadig mer tilgjengelig, parallelt med at elevenes og studentenes digitale kompetanse øker. Dette er fenomener som begge skulle tilsi økt bruk av korpus. Videre er det en sterk bevegelse for å åpne opp tilgang til forskningsdata og –ressurser (Open Access) og denne bølgen kan også brukes til å åpne opp forskningsressurser, herunder korpus, for læringsformål. Til syvende og sist er spørsmålet om bruk av korpus i klasserommet imidlertid avhengig av en tettere dialog mellom lærere og korpuslingvister, både i lærerutdanningen og gjennom faglig oppdatering underveis i yrkeslivet.

*1.      Introduksjon*

Lærebøker med prefabrikerte tekster som repeteres for fremmedspråkelever år etter år kan legge en demper på utforskergleden hos selv den mest engasjerte elev. Med korpusbasert undervisning får elever og studenter i fremtidens klasserom bedre tilgang til naturlig språklig råmateriale, som de kan utforske fritt eller ut fra spesifikke grammatiske og pedagogiske søkekriterier. Dette er en frisk og på mange måter mer naturlig tilnærming til fremmedspråket enn den tradisjonelle læreboken.

Norske elever og studenter har allerede svært høy digital kompetanse og en unik tilgang til datamaskiner og Internett. Men i hvilken grad blir digitale læremidler utnyttet i norske klasserom, og hvordan kan vi bruke dem mer og bedre i fremtiden?

I denne artikkelen skal vi ta for oss korpus, en type digitale språkressurser som spiller en viktig rolle i moderne lingvistikk, og som kan gi rask og enkel tilgang til autentisk språkmateriale i klasserommet. Spesielt for elever og studenter som har fremmedspråk som fransk, spansk og tysk kan korpus være viktig, fordi de møter disse språkene lite i hverdagen, i motsetning til engelsk som de møter kontinuerlig i filmer, bøker og sosiale medier.

I det følgende skal vi gi en del eksempler på hvordan korpus kan integreres i undervisningen og diskutere mulighetene for bruk av korpus i fremtidens klasserom. Vi vil snakke om både elever og studenter. Med *elever* mener vi elever i fremmedspråk (med særskilt vekt på 2. fremmedspråk som fransk, spansk og tysk, osv.) i den videregående skolen, mens vi med *studenter* mener studenter ved universitet og høyskoler på bachelor-nivå. Mye av det vi sier vil være gyldig i større eller mindre grad for begge gruppene, men der hvor det er hensiktsmessig skiller vi mellom dem.

Artikkelen er strukturert som følger: I del 2 gir vi en kort innføring i korpus og korpuslingvistikk. I del 3 gjør vi rede ulike problemstillinger knyttet til bruk av korpus i fremmedspråkundervisningen, som forskjellige typer bruk og forholdet til ulike pedagogiske paradigmer. Denne seksjonen gir også en oversikt over eksisterende forskningslitteratur på feltet. I del 4 beskriver vi to norske korpus som er spesielt egnet for bruk i fremmedspråkundervisningen, mens del 5 diskuterer hvorfor og hvordan en kan bruke korpus i klasserommet. Del 6 avslutter med en diskusjon av fremtidsperspektivene for bruk av korpus i norske klasserom.

*2.      Hva er korpus og korpuslingvistikk?*

De siste årene har lingvistikken blitt preget av en stadig økende bruk av digitale tekstsamlinger som kalles *korpus* (se McEnery & Hardie, 2012 som gir en god innføring). Korpus er databaser av digital, maskinlesbar tekst, ofte av et så stort omfang at det ville være umulig å gå igjennom dem for hånd. Fordi teksten er maskinlesbar kan vi bruke datamaskiner til å gå gjennom teksten og finne frem til de fenomenene vi er interesserte i.

Et korpus kan være innrettet mot et spesielt område eller teksttype, som vitenskapsartikler eller SMS-meldinger, eller det kan bestå av mange ulike typer sjangrer (avisartikler, romaner, osv.). Det er uansett et mål at korpuset skal være mest mulig representativt for den utvalgte teksttypen, slik at det kan gi et riktig bilde av språkbruken på området. Korpus kan også dekke ett enkelt språk, eller flere, dersom man ønsker å gjøre kontrastive analyser av ulike språk. Slike korpus er spesielt interessante for undervisning i fremmedspråk, og vi skal komme tilbake til disse senere (seksjon 4). Når det gjelder flerspråklige korpus, kan man skille ut ytterligere undertyper. *Oversettelseskorpus* inneholder originaltekster på et kildespråk og oversettelser på et målspråk. *Parallellkorpus* består av originaltekster fra flere kildespråk sammen med oversettelser på ett eller flere målspråk. *Oslo Multilingual Corpus* er et eksempel på parallellkorpus, og vi skal beskrive dette korpuset nærmere i seksjon 4 (se også Johansson, 2000). Slike korpus kan *parallellstilles* på setnings- eller ordnivå, slik at man raskt kan se hvilke enheter som svarer til hverandre på kilde- og målspråket. *Sammenlignbare* (’comparable’) korpus består av tekster som er på forskjellige språk, men av samme sjanger eller teksttype fra samme tidsperiode, og som derfor er sammenlignbare på tvers av språkene. *KIAP-*korpuset ved Universitetet i Bergen (Fløttum, Dahl & Kinn, 2006) består av vitenskapsartikler fra fransk, engelsk og norsk og er et godt eksempel på et sammenlignbart korpus. Noen korpus kan også inneholde en blanding av parallell- og sammenlignbare korpus.

*Referansekorpus* er store korpus som inneholder tekster fra ulike sjangrer og teksttyper, og gir dermed et godt bilde av språkbruken i det aktuelle språket. Det er et problem at mange språk mangler slike korpus, eller at de ikke er åpent tilgjengelige for allmennheten. For norsk bokmål kan prosjektet *Leksikografisk bokmålskorpus* (LBK), ledet av professor Ruth Vatvedt Fjeld (UiO), nevnes. Per august 2012 inneholdt korpuset 70 millioner ord, og det er stadig under oppbygging. Korpuset er fritt tilgjengelig for språkforskning, men man må søke om brukerkonto med passord og brukernavn for å bruke det. *British National Corpus* (BNC) er et tilsvarende eksempel for britisk engelsk, og tilbyr en rekke ulike søkefunksjonaliteter, både med og uten innlogging. For undervisningsformål kan det være praktisk med korpus der man slipper innlogging.

Man skiller også mellom skriftlige korpus og talekorpus, og i det siste er det også utviklet såkalte multimodale korpus, som gjerne inneholder video-opptak. Slike korpus kan være nyttige for å studere for eksempel samspillet mellom talespråk og bruk av gester.

I tillegg til rene korpusdata, kan korpuset forsynes med ytterligere informasjon som kan lette bruken. Slik tilleggsinformasjon – som også er maskinlesbar – kalles ofte *annotasjoner*. Annotasjoner kan gi informasjon om ordklassetilhørighet, syntaktiske relasjoner og andre fenomener som kan være nyttige for lingvistisk analyse. Spesielt er korpus som er tagget med ordklasseinformasjon, nyttig for fremmedspråkelever og studenter. Figur 1 viser et eksempel på en såkalt trebank, et korpus som er forsynt med informasjon om grammatiske og syntaktiske relasjoner, en avansert form for korpusannotasjon. Denne trebanken er en del av prosjektet INESS ved Universitetet i Bergen, en plattform som gir tilgang til trebanker for en rekke språk (se Rosén, De Smedt, Meurer & Dyvik, 2012 for en nærmere beskrivelse). Som en del av prosjektet bygger man også en trebank for norsk bokmål basert på modellen *Lexical Functional Grammar* (LFG), som dette utsnittet er hentet fra. Figuren viser hvordan en setning er forsynt med informasjon om bl.a. setningsrelasjoner og ordklasseinformasjon. I fremmedspråkundervisningen vil slike korpus være nyttige i innlæringen av grammatiske regler og for å trene opp ferdigheter i gjenkjenning av ordklasser og setningsanalyse.

**Figur 1 Trebanken INESS. Korpus annotert med syntaktisk informasjon.**



Korpus brukes ofte sammen med forskjellige dataverktøy. Blant de mest grunnleggende finner vi verktøy for å beregne *frekvens*, dvs. hvor hyppig et ord er brukt i det aktuelle korpuset. Et annet viktig verktøy er *konkordansverktøy*, som lar degsøke på nøkkelord og få opp konteksten til ordet du er interessert i. Dermed kan vi finne ut mye om hvordan ordet faktisk brukes. I Figur 2 ser vi et eksempel på konkordanser i såkalt KWIC-format (*Key Word In Context*), der man på forhånd kan spesifisere hvor stor kontekst man ønsker. Den store fordelen med konkordanser er at man kan fokusere på et bestemt fenomen ved å isolere relevante forekomster. Dermed blir mønstrene ekstra tydelige, som i dette eksempelet som viser konkordanser av preposisjonen *contre* fra et fransk delkorpus av KIAP-korpuset.

Når man skal studere preposisjoner kan et korpus være svært nyttig. De er notorisk vanskelige å få grep om for en fremmedspråkelev eller student, fordi oversettelsen i svært stor grad avhenger av konteksten preposisjonen opptrer i, spesielt for preposisjoner som selekteres av verb. Det er altså ikke mulig å gi én norsk hoved-preposisjon som oversettelse for hver L2-preposisjon. Det vil også ha begrenset verdi å gi en liste med ulike norske preposisjoner for hver enkelt L2-preposisjon, fordi preposisjoner i en slik liste er løsrevet fra konteksten. Siden kontekstuell informasjon er så avgjørende i slike tilfeller, er empirisk evidens hentet fra korpus svært verdifull.

**Figur 2 KWIC-konkordanser fra korpuset *KIAP***



*3. Bakgrunn: om bruk av korpus i fremmedspråkundervisningen*

Som nevnt tidligere er det stor interesse blant korpuslingvister for pedagogiske applikasjoner. Det eksisterer etter hvert også en nokså omfattende litteratur om emnet, og vi skal i denne seksjonen gi en kort oversikt over denne, der vi konsentrerer oss om ulike typer bruk av korpus, forholdet til forskjellige pedagogiske paradigmer og evaluering av fordeler og ulemper med bruk av korpus. Den eksisterende litteraturen finnes i hovedsak på engelsk, men Graedler, Egan & Nacey (2011) og Hasselgård (2014) er viktige bidrag innen norsk forskningslitteratur, og gir en god oversikt over en rekke korpus som kan anvendes i fremmedspråkundervisningen.

**3.1 Ulike typer bruk av korpus i fremmedspråkundervisningen**

Det er relativt uvanlig å bruke korpus i undervisningen, spesielt i grunnskolen og videregående utdanning. Heller ikke på universitetsnivå brukes korpus i noen særlig grad i fremmedspråkundervisningen, verken i Norge eller internasjonalt. Derimot brukes korpuslingvistikk i spesialiserte kurs innenfor allmennlingvistikk og spesielt datalingvistikk (Granath 2009, s. 48). Altså får bare en liten del av fremmedspråkelevene mulighet til å prøve ut korpuslingvistiske metoder på egenhånd.

Når det gjelder eksisterende bruk, er det vanlig å dele den i to kategorier (Leech, 1997), *indirekte* bruk, dvs. som referansemateriale for læreren, og *direkte* bruk av korpus i undervisningen. I tillegg er det i de siste årene dukket opp en tredje kategori (McEnery & Xiao, 2010), nemlig utviklingen av nye korpus spesifikt for læringsformål, for eksempel basert på materiale fra innlærerspråk (Graedler, Egan & Nacey 2011: 109-110). Disse er nyttige blant annet fordi de gir oss data som kan gi innsikt i mentale prosesser forbundet med språklæring.

I følge McEnery og Xiao (Ibid.) har den indirekte tilnærmingen vært klart dominerende, dels fordi lærere ikke alltid har den nødvendige forkunnskapen for å ta i bruk korpus, og dels fordi det ikke i tilstrekkelig grad er blitt utviklet korpusressurser som kan brukes til pedagogiske formål. I tillegg legger ofte lærebøker og læreplaner begrensninger på bruken av korpus og også på andre typer av digitale hjelpemidler.

Vi skal nå se nærmere på ulike former for indirekte og direkte bruk av korpus i klasserommet, med hovedvekt på direkte bruk.

*Indirekte bruk*

Indirekte bruk omfatter en rekke tema, som hovedsakelig har til felles at de fungerer som referansemateriale for læreren. En viktig funksjon vil derfor være at de setter læreren i stand til å gi gode, opplyste svar. Som fremmedspråklærer kan man ofte oppleve at elever spør om ting som ikke er dekket i læreboken, eller man kan komme over tvilstilfeller ved retting av prøver og oppgaver. Ofte kan læreren da føle behov for å rådføre seg med en morsmålsbruker, men det kan være upraktisk, og i utgangspunktet er heller ikke morsmålsbrukeres intuisjoner noen sikker kilde til kunnskap. I slike tilfeller kan derimot et korpus-søk gi et informert svar, og vil kunne gi verdifull empirisk innsikt i hvordan uttrykket faktisk brukes. Bruk av korpus er også mer troverdig og kvalitetssikret enn Google-søk, som mange bruker i dag. Et godt korpus vil alltid være bedre enn et nettsøk, ettersom man ikke kan vite noe særlig om kvaliteten på nettsidene man får treff på. Et unntak kan være svært sjeldne ord, som neppe er representert i korpus av begrenset størrelse, men som heller ikke vil være særlig relevante for elever og studenter på nivået det her er snakk om.

I dag brukes korpus i stor grad i utviklingen av referanseverk som referansegrammatikker og ordbøker, men også for å lage lærebøker og språkkurs. I tillegg kan korpusdata brukes i utviklingen av språktester. Videre kan korpusverktøy gi informasjon som kan brukes i utviklingen av pensum og undervisningsopplegg, som for eksempel frekvensanalyser. Dersom et bestemt fenomen er høyfrekvent vil det gi god mening å bruke tilsvarende mye tid på det i undervisningen, og pensumstoffet må også tilpasses. Et lavfrekvent fenomen trenger derimot ikke å prioriteres så høyt. Til en viss grad følges jo dette prinsippet også i dag, for eksempel ved undervisning i ulike verbtider, hvor det undervises lite eller ingenting i de lavfrekvente verbtidene, mens man bruker mye tid på de hyppigst brukte tidene. Korpusdata vil likevel kunne gi oss et mer solid fundament for slike vurderinger. Det må imidlertid understrekes at frekvensdata selvsagt må suppleres med andre pedagogiske hensyn, som for eksempel å fange opp fenomener knyttet til transferens fra L1, kontrast mellom L1 og L2 osv.

*Direkte bruk*

Direkte bruk dreier seg om undervisningsmetoder hvor elevene eller studentene selv får bruke korpus. Det finnes ulike typer av direkte bruk av korpus i fremmedspråkundervisningen. Først og fremst gir korpus tilgang til autentiske eksempler. Læreren kan på forhånd velge ut et sett med flere eksempler som illustrerer et bestemt fenomen. På denne måten kan man øve opp eleven eller studentens evne til å bite seg merke i språklige fenomener. Dette vil være en viktig kunnskap å ta med seg videre i språkstudiene. Forskningen har vist en rekke andre fordeler ved bruk av autentiske eksempler, som at de kan gjøre fremmedspråkundervisningen mer levende (Granath, 2009), og får undervisningsmaterialet til å fremstå som mer relevant for eleven (Bernardini, 2004).

Etter hvert som elevene eller studentene blir vant med analysen av korpusdata, kan de få tilgang til hele korpus, og må så sortere og velge ut den informasjonen som er mest relevant. Dette vil avhenge av nivået elevene eller studentene befinner seg på og lærerens egen kompetanse til å veilede dem. Det kan imidlertid innvendes at autentiske eksempler kan bli for vanskelige for elever i videregående skole og enkelte studenter, og at det er en lang tradisjon innenfor lingvistikk og fremmedspråkundervisning for bruk av fiktive eller tilpassede eksempler, selv om tendensen generelt nok går mot en stadig økende bruk av autentiske eksempler. Bruk av autentiske eksempler fra korpus kan nok være utfordrende for elever med kun elementære ferdigheter i L2, og det er da viktig at lærerne har den nødvendige opplæring til å kunne veilede, oversette og forenkle eksempler når dette er nødvendig.

Granath (2009, s. 60) hevder på sin side at det er en sammenheng mellom språklige ferdigheter og ferdigheter i korpuslingvistikk. Elever og studenter som i utgangspunktet har beskjedne ferdigheter i fremmedspråket, er som regel de samme som finner korpuslingvistiske metoder vanskelige. Det er antakelig en tilsvarende sammenheng for elever og studenter som har gode dataferdigheter i utgangspunktet. For disse kan korpusbaserte metoder brukes som en ekstra stimulus, og som virkemiddel for forsknings- og nysgjerrighetsdreven læring. Også elever og studenter med et mindre avansert ferdighetsnivå (*intermediate students*) vil kunne ha utbytte av korpusbaserte metoder, men vil da kreve en større grad av veiledning (Granath 2009, s. 63). Det kan derfor være en risiko for at de som har størst behov, vil få minst utbytte av korpusbruk, Granath (Ibid.) observerte imidlertid også at korpusundervisning på data-lab hadde en generell positiv effekt for språklæring, fordi det ga studentene i undersøkelsen større innsikt i nytten av å kunne grammatiske regler.

Bruk av korpusdata gir også elever og studenter mulighet til å reflektere kritisk over språkbruk (Aijmer 2009, s. 4). På et metanivå kan bruk av korpus også bidra til å gi bedre innsikt i ens eget morsmål, fordi det gir innsikt i hvor upålitelige intuisjoner kan være, og hvordan man kan verifisere eller avkrefte dem ved hjelp av empiriske data i et korpus (Bernardini 2004, s. 30).

Vi vil komme tilbake til andre konkrete applikasjoner av korpus i del 5.

**3.2 Pedagogiske paradigmer**

I en stor del av litteraturen knyttes bruken av korpus som læringsverktøy til spesifikke pedagogiske paradigmer; *data-driven learning* (se for eksempel Boulton, 2010), *discovery learning* osv. Felles for disse er en induktiv og nysgjerrighetsbasert tilnærming til læring, som i stor grad er student- og elevorientert (Ebeling, 2009, s. 81), og gir eleven muligheten til å ta ansvar for egen læring. Korpusbaserte metoder er også til en viss grad anti-autoritære, fordi det lar læreren fremtre som en veileder i en felles utforsking av korpus snarere enn mesteren som besitter det eneste riktige svaret:

[…] *learning as discovery* […] encourages learners to follow their own interests whilst providing them with opportunities to develop their capacities and competences so that their searches become better focused, their interpretation of results more precise, their understanding of corpus use and their language awareness sharper. This may be confusing at first, as learners are asked to abandon deeply rooted norms of classroom behavior, but soon becomes liberating for both teachers (who can stop pretending to be sources of absolute and limitless knowledge) and learners (who start to see themselves as active participants in the teaching-learning process). Bernardini (2004, s. 23)

Bernardini (2004, s. 16) forbinder også korpusbaserte metoder med det hun kaller ‘serendipitous learning’ (‘læring gjennom lykketreff’), fordi det å forholde seg til relativt omfattende autentisk språkmateriale kan gi tilgang til helt nye innsikter. Læringsprosessen kan da sammenliknes med en jakt etter språklige fenomener og regelmessigheter, der man plutselig kommer over noe uventet som kan gi ny innsikt.

Chambers (2010) påpeker at fremmedspråkelever kan ha særlig nytte av parallellkorpus, fordi disse er egnet til å belyse språklige fenomener kontrastivt. Korpus, og kanskje spesielt flerspråklige korpus og parallellkorpus gir en unik mulighet til å sammenlikne og kontrastere språklige fenomener på tvers av språkene. Det gir også elever og studenter anledning til å praktisere sin teoretiske kunnskap i grammatikk gjennom nysgjerrighetsbasert utforsking av korpus, på jakt etter regelmessigheter eller unntak som kan forklares med grunnlag i det grammatiske teoriapparatet de allerede har tilegnet seg.

Etter vårt syn er korpus et nyttig undervisningsverktøy dersom man ønsker å anvende slike læringsparadigmer, men det er fullt mulig å ha stort utbytte av korpus selv om man velger andre pedagogiske tilnærminger. Mer spesifikt mener vi at også en deduktiv, regelbasert undervisningsmetode kan ha god nytte av korpus (se også seksjon 5). Det er imidlertid fullt mulig å kombinere både induktive og deduktive metoder med en nysgjerrighetsbasert utforsking av data som læringsverktøy, og korpus kan ha sin funksjon innenfor begge paradigmer, selv om det tradisjonelt har blitt forbundet med induktive læringsmodeller.

**3.3 Fordeler og ulemper ved bruk av korpus i fremmedspråkundervisningen**

Når det er så mange tilsynelatende fordeler ved bruk av korpus, hva er så de potensielle hindringene for å ta det i bruk? Mangel på tid er selvsagt en faktor som begrenser muligheten til å ta i bruk korpus i undervisningen (Granath, 2009). Det er derfor viktig at bruk av korpus kobles til det aktuelle undervisningsmaterialet i grammatikk, og gjøres til en integrert del av undervisningen. Det skal heller ikke underslås at det trengs en del tid for å sette seg ordentlig inn i korpuslingvistikken, både for lærer og elever/studenter, for at resultatet skal bli vellykket. I tillegg kreves det ofte et visst nivå av grammatiske kunnskaper for at korpusbruken skal bli vellykket. Dersom eleven/studenten selv skal gå gjennom lister med autentiske eksempler hentet fra et korpus krever jo dette at hun/han har et minimum av innsikt i det grammatiske fenomenet som skal studeres. Det er derfor viktig med et tett samspill mellom bruk av korpus til observasjon av empiriske fenomener/induktiv læring og innlæring av grammatiske regler/deduksjon (vi kommer tilbake til dette i seksjon 5).

Mangel på teknisk utstyr er i dag ofte ikke noe problem i videregående skole, men i høyere utdanning kan man ikke nødvendigvis forvente at studentene har med seg bærbar datamaskin. Noen korpus vil bare kunne brukes i spesialiserte data-laber, men det er ønskelig at ressursen også kan lastes ned på elevens egen bærbare PC, slik at eleven kan jobbe med korpus-metoder hjemme. De fleste korpus vil imidlertid i dag ha nettbaserte grensesnitt som gjør at man kan jobbe med korpus hvor som helst, men kanskje trenger man en egen konto med innloggingsdetaljer og lisens, som i noen tilfeller koster penger. Tilgang til ressurser som kan brukes også utenfor klasserommet eller data-laben vil utvilsomt styrke elevens selvstendighet i møte med korpusressursene og vil kunne gi anledning til å supplere veiledet arbeid med egeninnsats.

For mange korpus er ikke grensesnittet tilpasset vanlige brukere, noe som kan utgjøre en hindring for mange elever og studenter. Dette gjelder både brukervennlighet og rent estetiske forhold. Ungdommer bruker daglig sosiale medier som Facebook og Instagram, som har svært enkle, tiltalende og brukervennlige grensesnitt. De fleste korpus kan umulig konkurrere med estetikken og brukervennligheten som slike nettsteder har, og dette kan utgjøre en hindring for at de tas i bruk. For mange korpus kan brukervennligheten med fordel forbedres før de kan tas i bruk i klasserommet. Det vil også være fordelaktig å utvikle pedagogiske verktøy eller grensesnitt som kan leveres sammen med korpuset.

Gruppestørrelse vil ofte være en utfordring, og fordi opplæring i bruk av korpus krever tett oppfølging fra lærerens side, kan det være ønskelig med små grupper. Spesielt elever og studenter med teknofobi og tallvegring vil kunne ha motstand mot å bruke korpus og forholde seg til store datasett og vil kunne ha utbytte av korpusundervisning i mindre grupper.

*4.      Eksempler på norske korpus:* Norwegian-Spanish Parallel Corpus *og* Oslo Multilingual Corpus

I dag er det nokså vanlig at korpus er passord-beskyttet og at man må søke eieren av det aktuelle korpuset om tilgang, gjerne med en begrunnelse og redegjørelse for bruken. Dette henger ofte sammen med hva som var hensikten med korpuset da det ble laget og tekstene ble samlet. Tekstene som inngår i et korpus er ofte beskyttet av åndsverksloven og forlag, forfattere osv. må således gi tillatelse til at tekstene inngår i et korpus. De som da inngår avtalen formulerer den gjerne slik at den dekker deres behov i det aktuelle prosjektet korpuset lages for. Når man senere for eksempel ønsker å bruke et gitt korpus i klasserommet er det ikke uvanlig at slike avtaler kan stå i veien for det. *Oslo Multilingual Corpus* (OMC), som vi bruker i denne artikkelen, er et eksempel på dette. OMC er bare tilgjengelig for studenter og forskere ved Universitetet i Oslo og Universitetet i Bergen, de to institusjonene som var med på å bygge opp korpuset. Vi har likevel valgt å bruke nettopp dette korpuset her på grunn av at det inneholder tekster som er oversatt til eller fra norsk, og fordi det på en god måte viser mulighetene et parallellkorpus gir i undervisningen. Vi håper og tror at det i fremtiden vil bli bedre tilgang på den typen korpus i den videregående skolen, både for lærere og elever. I *The Open Parallel corpus* (OPUS) (<http://opus.lingfil.uu.se/index.php>) finnes det imidlertid allerede en rekke andre parallellkorpus som kan benyttes uten å måtte søke om tilgang. Graedler, Egan & Nacey (2011) og Hasselgård (2014) gir forøvrig også gode oversikter over noen flere aktuelle korpus som kan være til nytte i en undervisningssammenheng.

**4.1 *The Norwegian-Spanish Parallel Corpus* (NSPC)**

Korpuset som er brukt som grunnlag for de spanske illustrasjonseksemplene er *The Norwegian-Spanish Parallel Corpus* som er kompilert ved Universitetet i Bergen (Hareide & Hofland, 2012). Dette er et oversettelseskorpus som inneholder enveisoversettelser fra norsk til spansk, og det består av 2,12 millioner spanske ord og 1,97 millioner norske, fra 31 tekstpar (Ibid., s. 76). Tekstene som er oversatt er moderne norske tekster, oversatt til spansk og publisert mellom år 2000 og 2009. Korpuset inneholder både skjønnlitterære tekster og tekster fra andre sjangrer. Det er lagt vekt på at det ikke skal være fagspråk, slik at det er tilgjengelig og lesbart for allmennheten (Ibid, s. 75). Hver enkelt tekst er klassifisert på grunnlag av sjanger, forfatterens kjønn, og oversetterens kjønn og morsmål. Korpuset skal tagges for ordklasser, som vil si at hvert ords ordklasse blir registrert og søkbart. Dette innebærer at korpuset etterhvert vil få langt flere ulike søkemuligheter, og vil i fremtiden bli enda bedre egnet som empirisk materiale for undervisning av en rekke ulike språklige fenomener.

**4.2 *Oslo Multilingual Corpus* (OMC)**

*Oslo Multilingual Corpus* (OMC) (1999-2008) (se https://www.hf.uio.no/ilos/tjenester/kunnskap/sprak/omc/) er et resultat av prosjektet *Språk i kontrast - SPRIK* ved Universitetet i Oslo (UiO) (se http://www.hf.uio.no/ilos/forskning/prosjekter/sprik/). SPRIK, som ble ledet av Stig Johansson og Cathrine Fabricius-Hansen, var et flerfaglig og tverrinstitusjonelt forskingsprosjekt som fokuserte på kontrastive språkstudier (engelsk, fransk, norsk, tysk). Prosjektet hadde som mål å styrke norsk språkforskning innenfor semantikk/pragmatikk, kontrastiv lingvistikk og stilistikk samt lingvistisk orientert translatologi. OMC ble utviklet av forskere fra Universitetet i Oslo og Universitetet i Bergen (se http://www.hf.uio.no/ilos/english/services/omc/team/) og er i dag tilgjengelig for forskere og studenter ved disse to universitetene. I fremtiden vil trolig tilgangen på slike korpus bli bedre, også for personer som ikke er tilknyttet forskningsinstitusjonene.

OMC er en samling tekstkorpus som består av originaltekster og oversettelser fra flere språk. De ulike underkorpusene skiller seg ved at de inneholder ulike språk eller ulike kombinasjoner av språk. OMCs forskjellige subkorpus kan deles inn i to hovedtyper av flerspråklig korpus: parallellkorpus og oversettelseskorpus.

Med parallellkorpus menes det i SPRIK en samling tekster som inneholder både originaltekster og oversettelser fra to eller flere språk. Et eksempel på parallellkorpus i OMC er delkorpuset *Fransk-norsk parallellkorpus* (*FNPC*) som vi beskriver under.

Med oversettelseskorpus menes det i SPRIK en samling av tekster som inneholder originaltekster fra ett språk med oversettelser til ett eller flere språk, med andre ord bare ett språk representert med originaltekster. Underkorpuset *Norsk-fransk-tysk oversettelseskorpus* (*No-Fr-Ge*) i OMC er et eksempel på dette.

I det følgende skal vi konsentrer oss om den franskspråklige delen av OMC, det fransk-norske parallellkorpuset (FNPC) som inneholder originaltekster og oversettelser (fransk-norsk, norsk-fransk). Korpuset består av en skjønnlitterær del (FNPC/Fiction) og en del med sakprosa (FNPC/Non-fiction). FNPC/Fiction omfatter seks franske originaltekster med oversettelser til norsk og fem norske originaltekster med oversettelser til fransk, mens FNPC/Non-fiction inkluderer 10 originale tekstutdrag fra hvert språk (med oversettelser), totalt ca. 864 600 ord. I tillegg inneholder det norsk-fransk-tyske oversettelseskorpuset (No-Fr-Ge) ca. 1 680 000 ord fra åtte norske skjønnlitterære tekster og deres oversettelser til fransk og tysk.

Subkorpusene FNPC og No-Fr-Ge er parallellstilt på setningsnivå, så i disse korpusene må man identifisere oversettelser av enkeltord manuelt. I OMC finnes det også subkorpus hvor parallellstillingen er gjort på ordnivå, noe som gjør det mye lettere å arbeide med tekstene i undervisningssammenheng. Det kreves imidlertid mye arbeid for å parallellstille alle tekstene på ordnivå.

*5.      Hvorfor og hvordan bruke korpus i undervisningen*

**5.1 Motivasjon for korpusbruk i fremmedspråkundervisningen**

Som vi har sett i seksjon 2 har bruk av korpus innenfor språkundervisning sin bakgrunn i to hovedkvaliteter en slik søkbar tekstsamling har, nemlig autentiske eksempler og frekvens (McEnery og Xiao 2010, s. 373). Autentisk språk er tekst og tale produsert naturlig i ulike sammenhenger, uavhengig av pedagogiske formål. Det er vesentlig at elever helt fra starten av blir eksponert for større eller mindre utdrag av autentisk L2, både fordi det er utopisk å skulle tilegne seg et fremmedspråk uten å se og høre hvordan dette språket fortoner seg i virkeligheten, og fordi at de relativt abstrakte strukturene elevene må få grep om i grammatikken bør få noen "knagger å henge på". Imidlertid er det notorisk utfordrende å finne enkle nok eksempler som skal illustrere et grammatisk eller leksikalsk poeng, hvis de også skal være autentiske. Med elektroniske korpus blir et søk etter gode og relevante eksempler gjort effektivt, slik at man får ivaretatt begge hensyn: autentisk språk og enkle, formålstjenlige eksempler.

Det er i hovedsak to bruksområder en fremmedspråklærer kan ha i tankene når hun/han benytter seg av et korpus: undervisningsforberedelser eller direkte bruk i klasseromsundervisningen. Disse to dimensjonene er selvsagt ikke helt uavhengige av hverandre.

Som ressurs i klasseromsundervisningen er et korpus et svært fleksibelt redskap, i den forstand at det ikke legger begrensninger på hvilken del av språket man kan belyse (leksikalsk, morfologisk, syntaktisk, sosiolingvistisk), eller hvilken pedagogisk tilnærmingsmetode man ønsker å ta i bruk (for eksempel tradisjonell grammatikk/oversettelsesmetode vs. naturlig/induksjonsbasert metode, se også seksjon 3).

Uansett hvilken vinkling man velger å ha i undervisningen, vil læreren måtte forholde seg til Utdanningsdirektoratets “Læreplan i fremmedspråk” (Utdanningsdirektoratet 2013). I læreplanen er det spesifisert tre hovedområder, hvor *Kommunikasjon* og *Språk, kultur og samfunn* er spesielt relevante for det vi snakker om her. “Kommunikasjon” omfatter bl.a.:

”…språklig repertoar – ordforråd, setningsbygning og tekstsammenheng – og spesifikke språklige ferdigheter som er nødvendig for å mestre ulike kommunikasjonssituasjoner. Nye medier og bruk av språket på tvers av fag og emner inngår også i dette hovedområdet” (Ibid).

“Språk, kultur og samfunn” omfatter bla: ”Arbeid med ulike typer tekster og møte med kulturelle uttrykksformer fra målspråklandet…” (Ibid).

Korpusbasert undervisning gir oss muligheten til å kombinere flere av de nevnte målene. Det elektroniske korpuset representerer et nytt medium innen data, og kan inkorporeres i ulike fagdisipliner, både språk (som vi skal se), kulturkunnskap og matematikk (statistikk).

Når det gjelder de andre målene for fremmedspråkundervisning som er spesifisert i lærerplanen, finnes det forskjellige oppfatninger av hvilken pedagogisk tilnærming som er den mest formålstjenlige. Enkelte metoder, som f.eks. *the Natural Approach*, fremmer en induktiv metode for innlæring av regler, og begrenset bruk av L1 og oversettelse (Steinberg og Sciarini 2006, s. 150-151). Ønsket er at elevene skal tilegne seg L2 mest mulig likt måten de tilegnet seg sitt eget morsmål på. Det finnes klare innvendinger mot å begrense seg til en “naturlig” innlæring av språk i en så unaturlig kommunikasjonssituasjon som klasseromsundervisningen er; likevel har induktiv metode og glosetilegnelse på bakgrunn av kontekst heller enn oversettelse klare pedagogiske fordeler. Strukturer man får grep om på grunnlag av egen resonnering, sitter ofte bedre fast i minnet enn de man får direkte overført, ferdig fordøyet, fra en bok eller en lærer (se også seksjon 3).

**5.2  Eksempler på bruk av korpus i fremmedspråkundervisningen**

Ønsker man en induktiv tilnærming til undervisningen, vil korpuset forsyne læreren med en unik mulighet til å lage gode oppgaver. Det er mange ulike fenomener som kan belyses ut fra et slikt perspektiv. I det følgende gir vi noen eksempler på bruk av korpus i undervisningen.

*Ordklasser*

Ett av de mest sentrale temaene elever og studenter må settes inn i når de lærer et fremmedspråk, er ordklasser. I mange tilfeller vil tradisjonelle, semantiske kriterier for inndeling av ordklasser fungere dårlig når man skal identifisere dem. For eksempel, hvis det er slik at verb uttrykker handling, adjektiver peker på kvaliteter, og adverb kan vise til tid, hvordan kan det da ha seg at “(et) hopp”, “skjønnhet”, og “uke”, alle er substantiver? Et korpus-søk kan hjelpe oss å identifisere de *formelle* kriteriene for definisjonen av ordklasser, som vil inneholde langt færre unntak enn de semantiske, i tillegg til at de svært ofte er språkspesifikke, og dermed gir elevene og studentene et innblikk i hva som kjennetegner nettopp L2. De kan bli satt til å søke åpent på “substantiv” for eksempel, og dernest registrere hva et substantiv kan bøyes i, og hvilke ord det ofte opptrer sammen med. Da vil de, for mange språk, oppdage at substantiver kan opptre med artikler, og bøyes i tall.

*Vokabular*

For å øve inn gloser kan eleven/studenten settes til å gjøre korpus-søk og finne ut hvordan et ord opptrer i kontekst, hvilke typer ord det kan og må opptre sammen med, og hva slags semantisk innhold ordet bidrar med til konteksten. Denne informasjonen kan brukes til minst to ting: a) å indusere seg frem til ordets betydning, b) å lære kollokasjoner, altså hvilke faste/vanlige uttrykk ordet er del av. Når det gjelder parallellkorpus, som vi snakker om her, vil den første av de to nevnte oppgavene også berikes av tilgangen til oversettelsen av alle setningene et ord opptrer i. Slik vil korpuset gi et nyansert og fyldig bilde av hva slags betydning et ord tilfører konteksten, samt en liste med ulike ord det kan oversettes til, i en meningsfylt sammenheng. Mange elever og studenter blir forvirret av at ordbøker tilbyr et utvalg ulike norske ord for hvert ord på fremmedspråket, og fordi de mangler informasjon om hvordan ordet skal brukes i kontekst, velger de ofte feilaktige uttrykk som oversettelse i ulike sammenhenger. Med et korpus-søk unngår man dette problemet, samtidig som eleven/studenten må gjøre en betydelig egeninnsats for å komme frem til, og sortere, de ulike termene ett enkelt uttrykk kan oversettes med.

Et korpus-søk gir også informasjon om *kollokasjoner*, som er større, faste ordsamlinger, og disse er helt sentrale for produksjon og forståelse av et språk. Cowie forklarer: “native-like proficiency of a language depends crucially on knowledge of a stock of prefabricated units” (Cowie, sitert i McEnery og Xiao, 2010, s. 368). Slik informasjon får man i begrenset grad fra en vanlig ordbok. Et korpus-søk vil raskt avsløre om et ord til stadighet opptrer sammen med bestemte andre. Eksempelvis kan man gi studenten i oppgave å søke på det spanske ordet “toca”, som til vanlig betyr “berører” eller “spiller (instrument)”, og be dem legge merke til om det i noen sammenhenger får en annen betydning, og om det da står sammen med bestemte andre ord. Søket kan utføres som vist i Figur 4 og Figur 5.

**Figur 4 Søkespesifikasjonene for verbformen ’toca’ i NSPC**



**Figur 5 Resultater av søket på verbformen ’toca’**



Dette søket avslører blant annet at verbet “tocar”, når et objektspronomen er foranstilt, betyr “å være X sin tur”.

Det kan virke motiverende for elever og studenter å få jobbe slik med språklig råmateriale istedenfor ferdig fordøyet lærebokmateriale. Slik får de en smakebit på hvordan det er å være forsker, noe som ikke har vært vanlig i tradisjonell fremmedspråkundervisning.

*Grammatiske konstruksjoner*

Et parallellkorpus er dessuten spesielt verdifullt når man ønsker å belyse spesifikke grammatiske konstruksjoner, eller egne setningstyper, som for eksempel passiv. Passiv kan uttrykkes på flere svært ulike måter på spansk, og noen av disse inneholder ord som ikke har noen tilsvarende form i norske passivsetninger. Et eksempel er det refleksive pronomenet “se”, som i “Se venden libros”, som betyr “bøker blir solgt”. Vår erfaring har vært at elever blir forvirret av disse konstruksjonene, og spør til stadighet hva ordet “se” betyr. Mange sliter med å forstå at det ikke alltid er mulig å oversette hvert enkelt ord i en setning. Med et parallellkorpus vil de få tilgang til en stor mengde eksempler på samme type konstruksjon. Her kunne man tenke seg flere mulige oppgaver. Enten kunne man sette dem til å søke på norske passivkonstruksjoner, hvor søkeordene kunne være lemmaet “være” eller “bli” + partisipp. Treffene på et slikt søk ville utvilsomt gi dem spanske passivkonstruksjoner av flere forskjellige typer, og jo flere eksempler man får på samme struktur, jo lettere øyner man hva det er eksemplene har til felles, altså hva “se” faktisk uttrykker. En alternativ oppgave kunne være simpelthen å søke på det spanske pronomenet “se”, og registrere de tilfellene der setninger med dette pronomenet er oversatt med en passivsetning til norsk. De kunne få i oppgave å finne ut hva det er som skiller de spanske passivsetningene fra andre setninger der det nevnte pronomenet opptrer.

Det kan også finnes gode grunner til å ønske at elevene og studentene har en deduktiv tilnærming til enkelte tema innen fremmedspråket. På grunn av korpusets varierte søkemuligheter vil det også egne seg godt som materiale for en slik fremgangsmåte.

Vår erfaring som fremmedspråklærere, både i videregående skole og i høyere utdanning, er at elever og studenter ofte har sterk motstand mot å tilegne seg grammatisk terminologi. Mye av grunnen til dette, i tillegg til at ordene er fremmede for dem, er at de ikke ser poenget med det. Med et elektronisk korpus kan man illustrere hvor uvurderlig slik terminologi er, fordi man i korpus-søk kan selektere helt presis og konkret språklig informasjon, som man ikke ville kunne få på noen annen måte. Når man har som mål at elevene og studentene skal forstå poenget med å lære grammatisk terminologi, kan man ta i bruk deduktiv og induktiv metode i oppgavene. De kan få i oppgave å finne ett enkelt søkeuttrykk som gjør det mulig å søke etter uttrykk av følgende type: “estoy cantando” (“jeg synger”), “estábamos comiendo” (“vi spiste”), “estás viviendo” (“du lever/bor”), “están bailando” (“de danser”). Løsningen på oppgaven er å søke på lemmaet “estar” (som vil gi alle bøyningsformer av dette verbet i korpus), og *gerundium* av verbet som følger. Denne tilnærmingen kan man ha til all grammatisk terminologi, fordi terminologi til syvende og sist er generalisering over språklige regelmessigheter. Videre kan de få i oppgave å gå deduktivt til verks, og se hvilke korpustilfeller som stemmer overens med lærebokbeskrivelsen av det gjeldende uttrykket.

*Sosiolingvistiske tema*

Korpus er også nyttig for fremmedspråkundervisning som ikke tar for seg rent grammatiske emner. Man kan søke i korpus for å studere kulturelle særtrekk eller holdninger. Elevene eller studentene kan bli bedt om å søke på en politisk ladet eller relevant term, som “fattigdom”, “kvinne”, eller “mann”, og registrere hvilke ord de opptrer sammen med, og hvilke sammenhenger de er brukt i. Forskningsprosjektet “Poverty, Language, and Media” (Forskningsrådet 2013) ved Institutt for Fremmedspråk (UiB), ledet av Ana Beatriz Chiquito, er et godt eksempel på hvordan korpus kan brukes for å studere slike fenomener. Dette prosjektet, som bruker et korpus bestående av tekster produsert av og for massemedia i Colombia, Mexico, Argentina og Brasil, undersøker hvordan media i disse landene omtaler fattigdom, og dermed hvordan innholdet i dette begrepet formes hos publikum. Et slikt interdisiplinært perspektiv (lingvistikk og kulturstudier) kan med fordel inkorporeres i fremtidens fremmedspråkundervisning, og slik understreke at språk og kultur ikke er to uavhengige størrelser.

*Syntaktiske funksjoner og setningens oppbygning*

Etter hvert som elevene og studentene har fått på plass grunnleggende ferdigheter innen korpuslingvistikk, kan de få mer komplekse oppgaver der de ser på samspillet av flere fenomener samtidig. De kan for eksempel få i oppgave å observere samspill mellom verb og preposisjoner, som for det franske verbet *parler* (å snakke), og hva valget av preposisjon etter verbet har å si for hvilke komplementer verbet tar. De kan også få i oppgave å se om det finnes et én-til-én-forhold mellom preposisjonene etter *parler* og preposisjonene etter *snakke* i den norske oversettelsen. En slik kontrastiv oppgave vil gi dem innsikt i både preposisjonens påvirkning på tolkningen av verbet, og i preposisjonens ulike funksjoner i setningen. Læreren kan videre ved en setningsanalyse vise hvordan valget av preposisjon påvirker hvilke syntaktiske funksjoner setningens ledd får på grunnlag av preposisjonen(e) som følger etter *parler* (indirekte objekt, dativobjekt eller adverbial) og knytte dette opp mot betydningen til setningsleddene. En slik oppgave kan løses med OMC hvis man for eksempel søker på ’parle’, den bøyde formen av verbet *parler* (presens indikativ, 1. og 3. person):

**Figur 6 Søkegrensesnitt i OMC (FNPC/Fiction)**



Figur 7 viser et utdrag av resultatet fra treffene på dette søket:

**Figur 7 Utdrag av parallellstilte konkordanser for søk par ’parle’ i FNPC**

(Det er vi som har understreket enkelte preposisjoner og det pronominale uttrykket ’en’ i tilknytning til verbet på begge språkene).



Søkeresultatene viser hvordan preposisjonene kan angi forskjellige syntaktiske funksjoner; indirekte objekt (’de’), dativobjekt (’à’) og måtesadverbial (’avec’) og hvordan dette er oversatt til norsk. Det siste søkeresultatet viser også at det ikke alltid er en én-til-én korrespondens mellom språkene i bruken av preposisjoner. Vi ser at uttrykket ” … mais tout le monde ne **parle** que …” oversettes med ”… men folk **snakker** ikke **om** annet enn”. På norsk er det altså nødvendig å føye til preposisjonen *om* for å få en grammatisk korrekt setning. Dette er noe man ofte ser i parallellstilte tekster og er en utfordring når man jobber med å parallellstille tekster på ordnivå. Etter hvert som slike korpus forhåpentligvis blir mer tilgjengelige i fremtiden, vil de, i tillegg til å bli parallellstilt på ordnivå, også blir tagget for andre ting som for eksempel ordklasser, verbtider, modus, eller syntaktiske funksjoner, og dermed vil nytten i undervisningen bli enda større. Da vil man enkelt kunne finne bestemte syntaktiske konstruksjoner og sammenligne den med tilsvarende setninger i oversettelsen til eller fra fremmedspråket. Elevene eller studentene kan da finne store mengder setninger med samme syntaktiske struktur som den de ønsker å undersøke.

*6. Avsluttende bemerkninger: fremtidsperspektiver*

Allerede nå er det altså gode grunner til å ta i bruk korpusressurser i norske klasserom. Men hvordan ser fremtiden ut for denne typen læringsverktøy? For å si noe om dette må vi se på en del generelle trender knyttet til teknologi og digitale forskningsressurser.

For det første er digitale teknologier blitt mye mer tilgjengelig for norske skole-elever, og denne tendensen vil nok ikke svekkes. Vi har også fått en økt tilgjengelighet av forskningsressurser (tekstsamlinger og annet empirisk materiale og tilhørende dataverktøy) gjennom distribuerte forskningsinfrastrukturer og skyteknologi. De siste årene har det vært en stor satsing på internasjonale forskningsinfrastrukturer, som gjør det lettere å dele forskningsdata og ressurser som er utviklet fra disse. Innenfor språkvitenskap og korpuslingvistikk kan vi nevne det europeiske initiativet CLARIN (*Common Language Resources and Technology Infrastructure*, se www.clarin.eu) som lar europeiske forskere dele sine ressurser med hverandre. Beklageligvis inneholder disse plattformene bare i begrenset grad undervisningsmateriale, og det er sterkt ønskelig at det utvikles europeiske og internasjonale plattformer for deling av pedagogisk materiale, både for videregående og høyere utdanning.

Nye former for teknologi, som skyteknologi, dvs. at data og ressurser kan lagres på nettet og ikke trenger å ligge på fysiske datamaskiner, legger også til rette for at slik deling kan skje enklere. Ressursene vil dermed være enkelt tilgjengelige både hjemme og på skolen. Det haster å ta dette i bruk i skolen.

Parallelt med at myndighetene legger teknisk til rette for deling av ressurser, jobber enkeltpersoner og frivillige organisasjoner for åpne data og ressurser. I de siste årene har det også vært et økende ønske om å åpne opp forskningsdata, både blant forskere selv og bevilgende myndigheter. I Storbritannia er det for eksempel et krav fra Forskningsrådet (Research Councils UK) om at forskningsresultater finansiert av dem skal gjøres tilgjengelig via Open Access. Det er imidlertid på tide at denne tankegangen også anvendes når det gjelder å stille forskningsressurser til rådighet for pedagogisk bruk. I fremtiden bør bevilgende myndigheter også inkludere ’pedagogiske applikasjoner’ i søknadsskjemaene når forskere søker om midler. Bedre tilgjengelighet av pedagogiske applikasjoner basert på forskningsressurser som korpus vil også kunne gi et løft for elev- og studentdreven forskning, eller det som kalles ’citizen science’, ved å gjøre forskningsresultater lettere tilgjengelig for brukere i begynnerfasen. I forskningskonkurranser som *Google Science Fair* har man allerede sett en rekke eksempler på at ungdommer har utviklet interessante applikasjoner for kreftforskning, for eksempel, som blant annet har blitt mulig fordi data er lettere tilgjengelige. Innføring av forskningsbaserte ressurser tidlig i utdanningsløpet vil kunne bidra til mer av dette i fremtiden.

Dagens elever og studenter har en sterk digital kompetanse, og antagelig vil den fortsette å vokse. Dette gjør at skolen og lærerne må møte dem på den digitale arenaen. Det er imidlertid viktig å skille mellom ulike typer av digitale læringsaktiviteter og bruk av Internett til spill og sosiale medier, som kan fungere læringsfremmende, men som også kan være en oppmerksomhetstyv. Det er viktig at lærerne og skolen er i stand til å tilby seriøse læringsaktiviteter som gir god læringseffekt, og korpus er svært formålstjenlige i så måte. Men samtidig som ungdommen i stadig større grad lever digitale liv og dermed øker sin digitale kompetanse, er det en fare for at deres teknologiske kompetanse paradoksalt nok delvis svekkes gjennom den stadig økende bruken av smarttelefoner og *tablets*. Slike apparater kan ha brukervennlige og intuitive grensesnitt, men kan sies å være et tilbakeskritt når det gjelder å gi brukeren innsikt i teknologien og programmeringen som ligger bak. Bevegelsen *Lær kidsa koding* er en reaksjon på dette, og understreker at koding og programmering må regnes som en basisferdighet i dagens samfunn, og at skolen derfor burde bruke mer tid på dette. Korpusbasert undervisning kan selvsagt ikke i seg selv bøte på dette problemet, men vil likevel kunne gi innsikt i en type dataprogrammer som er ganske annerledes enn den typen spill og sosiale medier som mange ungdommer ofte oppsøker. Det vil også være rom for å legge inn undervisning i enkel koding, slik at man for eksempel kan lage sine egne søkekommandoer.

Men hvis digitale ressurser er på fremmarsj, hvordan ser så fremtiden ut for papirboken? Allerede i dag ser vi at digitale ressurser er i ferd med å utkonkurrere de tradisjonelle papirbaserte skolebøkene på visse områder. Det er selvsagt umulig å oppdatere en papirbasert lærebok like fort som digitale ressurser. Når det er sagt har papirboken klare fordeler når det gjelder læring, og det er vanskelig å se for seg fremmedspråkundervisningen uten papirboken i bunn. Vi ser det verken som ønskelig eller fruktbart å fjerne papirboken fra undervisningen. Likevel kan man se for seg løsninger der papirlæreverk i økende grad suppleres med digitale ressurser, og de fleste læreverk har allerede i dag en tilknyttet, nettbasert læringsplattform som det ville være enkelt å utvide med korpusbaserte øvelser. Digitale ressurser kan også kombineres med ’print-on-demand’-løsninger slik at lærebøkene kan oppdateres hyppigere. Bruk av korpus kan uansett integreres med både dagens og fremtidens lærebøker, og trenger ikke sees på som en helt ny metode.

Endelig tror vi at en innføring av korpus i fremmedspråkundervisningen må sees i sammenheng med en eventuell omlegging av fremmedspråkundervisningen på universitets- og høyskolenivå. Mye tyder nemlig på at dette nivået må revideres grunnleggende, blant annet på grunn av at det er stort frafall av studenter, og usikkerhet om hvordan filologutdannelsen kan brukes i yrkeslivet. Vi mener at korpuslingvistikk i større grad må integreres i undervisningen, både i form av en generell innføring som gir innsikt i bruk av korpus, og gjennom undervisning i hvordan korpus kan brukes når man selv skal ut i klasserommene. Korpuslingvistikk burde betraktes som en naturlig del av fremmedspråkundervisningen i høyere utdanning, og spesielt som en del av fremmedspråkdidaktikken. Det bør også gis tilbud om kurs i korpuslingvistikk som en del av lærernes etter- og videreutdanning.

Gjennom undervisningsreformer på dette nivået vil man også kunne nå det som er den kanskje viktigste aktøren for å få innført korpus i norske klasserom: læreren. Læreren spiller en avgjørende rolle dersom korpus skal kunne tas i bruk i undervisningen. Det forutsetter at læreren har en nødvendig innsikt i korpuslingvistiske metoder, men også at læreren er genuint interessert i korpusmetoder, og ser nytten og relevansen av dem. Men i dag er det fremdeles en for stor avstand mellom korpuslingvister og klasserommet. En tettere dialog mellom de to feltene er derfor avgjørende for at vi skal få fullt utbytte av den nye teknologien.

**Referanseliste**

Aijmer, K. (red.). (2009). *Corpora and Language Teaching*. Amsterdam:

John Benjamins Publishing Company.

Aijmer, K. (2009). Introduction. Corpora and language teaching. I Aijmer, K. (red.). *Corpora and Language Teaching*. Amsterdam: John Benjamins Publishing Company, s. 1-10.

Bernardini, S. (2004). Corpora in the classroom: An overview and some

reflections on future developments. I J.M. Sinclair (red.), *How to use corpora in language teaching.* Amsterdam: John Benjamins Publishing Company, s. 15-36.

Bjørke, C., Dypedahl, M. & Myklevold, G.-A. (2014). *Fremmedspråksdidaktikk*. Oslo:

Cappelen Damm.

Boulton, A. (2010). Data-Driven Learning: On Paper, In Practice. I T. Harris & M. Moreno Jaén (red.), *Corpus Linguistics in Language* *Teaching*. Bern: Peter Lang, s. 17-52.

Chambers, A. (2010). Contrastive Language Data: From Translation Studies to Language Learning and Teaching. I T. Harris & M. Moreno Jaén (red.), *Corpus Linguistics in Language* *Teaching*. Bern: Peter Lang, s. 99-118.

Dyndahl, P., Engen, T. O. , & Kulbrandstad, L. I. (red.) (2011). *Lærerutdanningsfag, forskning og forskerutdanning. Bidrag til kunnskapsområder in endring.* Vallset: Oplandske Bokforlag.

Ebeling, S. O. (2009). Oslo Interactive English. Corpus-driven exercises on the Web. I K. Aijmer (red.), *Corpora and Language Teaching*. Amsterdam: John Benjamins Publishing Company, s. 67-82.

Fløttum, K., Dahl, T., & Kinn, T. (2006). *Academic Voices – across languages and*

*disciplines.* Amsterdam: John Benjamins Publishing Company.

Graedler, A.-L., Egan, T. & Nacey, S. (2011). Korpus og korpusanalyse I språkdidaktisk forskning og praksis. I P. Dyndahl, T. O. Engen & L. I. Kulbrandstad (red.), *Lærerutdanningsfag, forskning og forskerutdanning. Bidrag til kunnskapsområder in endring.* Vallset: Oplandske Bokforlag, s. 101-127.

Granath, S. (2009). Who benefits from learning how to use corpora? I K. Aijmer (red.), *Corpora and Language Teaching*. Amsterdam: John Benjamins Publishing Company, s. 47-65.

Hareide, L. & Hofland, K. (2012). Compiling a Norwegian-Spanish parallel corpus. I M. P. Oakes, & M. Ji (red.). *Quantitative Methods in Corpus-based Translation Studies: A Practical Guide to Descriptive Translation Research.*Amsterdam: John Benjamins Publishing Company, s. 75-114.

Harris, T. & Moreno Jaén, M. (red.). (2010). *Corpus Linguistics in Language*

*Teaching*. Bern: Peter Lang.

Hasselgård, H. (2014). Å bruke korpus i språkundervisningen. I C. Bjørke, M.

Dypedahl &  G.-A. Myklevold (red.), *Fremmedspråksdidaktikk*. Oslo: Cappelen Damm, s. 139-156.

Johansson, S. (2000). Towards a multilingual corpus for contrastive analysis and

translation studies. *SPRIKreports*, 1.

McEnery, T. & Xiao, R. (2010). What corpora can offer in language teaching and

learning. *Handbook of Research in Second Language Teaching and Learning* 2, s. 364-380.

McEnery, T. & A. Hardie (2012). *Corpus Linguistics*. Cambridge: Cambridge

University Press.

Oakes, Michael P. & Ji, M. (red.). *Quantitative Methods in Corpus-based Translation Studies: A Practical Guide to Descriptive Translation Research.*Amsterdam: John Benjamins Publishing Company.

Rosén, V., De Smedt, K., Meurer, P. & Dyvik, H. (2012). An open infrastructure for

advanced treebanking. I J. Hajič, K. De Smedt, M. Tadić, & A. Branco (red.), META-RESEARCH Workshop on Advanced Treebanking at LREC2012, s. 22-29.

Römer, U. (2009). Corpus research and practice: What help do teachers need and what can we offer? I K. Aijmer (red.), *Corpora and Language Teaching*. Amsterdam: John Benjamins Publishing Company, s. 83-98.

Sinclair, J. M. (red.). (2004). *How to use corpora in language teaching.* Amsterdam: John Benjamins Publishing Company.

Steinberg, D. D. and N. Sciarini (2006). *Introduction to Psycholinguistics*, Pearson Education.

Utdanningsdirektoratet (2013). ”Læreplan i fremmedspråk.”

<<http://www.udir.no/kl06/FSP1-01/Hele/Hovedomraader/>> Hentet 28.01.2014.

**Korpus:**

Oslo Multilingual Corpus (1999-2008), Det humanistiske Fakultet, Universitetet i Oslo. <http://www.hf.uio.no/ilos/english/services/omc/>

The Norwegian-Spanish Parallel Corpus. [<http://metashare.nb.no/repository/browse/the-norwegian-spanish-parallel-corpus/6d4949c254e111e2b03a001708556d5a1ec93bda38654314b07ab03fa3ec37ba/](http://metashare.nb.no/repository/browse/the-norwegian-spanish-parallel-corpus/6d4949c254e111e2b03a001708556d5a1ec93bda38654314b07ab03fa3ec37ba/)>

The Open Parallel corpus (OPUS) <<http://opus.lingfil.uu.se/index.php>>