

Oslo Interactive English – om bruk av tekstkorpora i språkundervisning

Signe Oksefjell Ebeling

Institutt for litteratur, områdestudier og europeiske språk
Universitetet i Oslo
s.o.ebeling@ilos.uio.no



Hilde Hasselgård

Institutt for litteratur, områdestudier og europeiske språk
Universitetet i Oslo
hilde.hasselgard@ilos.uio.no
<http://folk.uio.no/hasselg>



Signe Oksefjell Ebeling er utdannet Dr.art. i engelsk språk og er prosjektleder for e-læringsprosjektene Oslo Interactive English og Trivial Corpus Pursuit ved Universitetet i Oslo. Hun er også tilknyttet prosjektet British Academic Written English Corpus ved Oxford Brookes University og har tidligere arbeidet med utviklingen av Oslo Multilingual Corpus.

Hilde Hasselgård er professor i engelsk språk ved Universitetet i Oslo. Hun har arbeidet mye med både engelskspråklige og flerspråklige korpora i sin forskning og undervisning og har sittet i styringsgruppen for Oslo Interactive English. Tidligere har hun vært med å tilrettelegge kurset Corpus Linguistics for fjernundervisning via internett.

Innledning

Engelskmiljøet ved Universitetet i Oslo har i mange år drevet korpusbasert språkforskning. Når det gjelder bruk av korpus til undervisning, har dette inntil relativt nylig kun blitt gjort på høyere nivåer; dvs. at studenter har brukt korpus til sine hovedoppgaver.

I 2004 fikk Institutt for britiske og amerikanske studier bevilget "støtte til utvikling av fleksible læringsformer" fra Universitetet i Oslo. Det var en forutsetning for tildeling av midler at opplegget skulle komme campus-studenter til gode, og ikke brukes kun til fjernundervisning. Instituttet erkjente at studentene hadde behov for mer praktisk språklig øvelse enn det vi hadde ressurser til å gi dem gjennom den tradisjonelle, lærerorganiserte undervisningen. Vi ønsket å gi lavere grads studenter mulighet til å lære å ta i bruk korpus samtidig som de oppøvde sine ferdigheter i engelsk språkføring og grammatikk. Parallelt med dette ønsket vi å skape en modell som ville oppmuntre studentene til å arbeide aktivt med språket og selv gjøre oppdagelser om

språket i bruk. Prosjektets mål var dermed å utvikle et korpusbasert opplegg for praktisk språklæring som samtidig gir innsikt i språklige strukturer og funksjoner.

Hovedtrekk i Oslo Interactive English

Oslo Interactive English (OIE) er et opplegg for språklæring som inneholder et engelskspråklig korpus, oppgaver i engelsk språkbruk og lenker til grammatikker, ordbøker og andre hjelpemidler som er tilgjengelig online. Alle oppgavene er utviklet spesielt til prosjektet, og de kan løses ved hjelp av søking i korpuset. Søkegrensesnittet til korpuset er inkorporert i OIE-grensesnittet, slik at studentene kan være på ett nettsted. Opplegget er helt autonomt, dvs. at det ikke er noen lærer som veileder eller retter oppgaver. Arbeidet med tilrettelegging av korpuset og utvikling av oppgaver og nettsted er utført av Signe Oksefjell Ebeling, som ble lønnet av prosjektet i halv stilling i et år.

Målgruppen for OIE er hovedsakelig studenter i sitt første år av engelskstudiet. Tilnærmingen til grammatikk og terminologien som er brukt, er derfor den samme som i lærebøkene som brukes på dette nivået (særlig Hasselgård m fl 1998 og Lysvåg 2005). Temaene som tas opp, gjenspeiler områder som mange studenter erfaringsmessig har problemer med på dette nivået. OIE er altså tenkt som et supplement til den ordinære undervisningen. De enkelte lærere oppfordres også til å bruke det aktivt i sine kurs, spesielt i begynnelsen av et nytt semester.

Oppgavetyper

For å få uerfarne studenter til å ta i bruk korpus, er det viktig at korpusbruken har et klart definert formål. OIE er derfor strukturert som en samling med oppgaver av typen "data-driven learning" (kfr. Tim Johns's hjemmeside). Hovedområdene i oppgavene er følgende:

- Adjectives & adverbs
- Nouns & verbs
- Function words
- Vocabulary
- Text
- English vs. Norwegian
- Miscellaneous

Innenfor hvert område er det flere undertemaer. Figur 1 viser undertemaene som finnes under "Nouns and Verbs".

Introduction: The exercises below focus on the uses and characteristics of the word classes noun and verb. Both nouns and verbs typically carry lexical content and their interaction is crucial for the syntax and meaning of a text. Their interaction is particularly apparent in Subject-Verb concord and collocations. These are topics that will be taken up in the exercises. Both nouns and verbs follow certain inflectional and derivational patterns and some of the exercises concentrate on problems related to the form of both nouns and verbs.

Aims: The exercises aim to make you familiar with some of the potential problems related to the use of nouns and verbs in English. By doing the exercises, you should be better equipped to use nouns and verbs according to the grammatical rules for standard English.

Exercises:

- Concord | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 |
- Countable vs. uncountable nouns | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 |
- -ing or to-infinitive? | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 |
- Modal auxiliaries | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 |
- Verb collocations | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 |
- Verb forms | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 |

Figur 1: Eksempel fra oppgaveoversikt i OIE; åpningsiden til "Nouns & Verbs"

Hvert undertema har et definert sett med oppgaver: *multiple choice* (1+2), *gapping* (3+4), *error correction* (5+6), og *open* (7+8). Under "multiple choice", får studenten et sett med alternative svar. Riktig svar kan finnes ved å utføre et søk i korpuset. Som regel vil det være råd om hvordan søkingen skal foregå. "Gapping"-oppgavene er innfyllingsoppgaver, som kan gjelde praktisk språkbruk, som i eksempel 1 (nedenfor) eller grammatisk analyse, som i eksempel 2. Setningene i disse oppgavene er (eventuelt i lett bearbeidet form) fra engelskspråklige korpora. Her er det ikke nødvendig å bruke OIE-korpuset for å løse oppgavene, men korpuset vil være til støtte. Når alle oppgavene er gjort, trykker studenten på "Submit" og får opp sin score, en lenke til fasit på oppgavene, samt en beskjed av typen

If your score was less than 100% you should consider the following:

- Read about verb + *-ing* or *to*-infinitive **here** and **here**
- Get corpus search hints **here**.

Eksempel 1:

[-ing or to-infinitive? #3]

Fill in the gaps with the *-ing* form or the *to*-infinitive of the verb in brackets. If both forms are possible, choose the one that has been most commonly used in the corpus.

(1) He listened without anything. (say)

Eksempel 2:

[Countable vs. uncountable nouns #3]

Decide whether it is the countable or uncountable variant of the nouns in question that has been used. Indicate your choice by using lower case **c** (countable) or **u** (uncountable) in the blanks.

Countable or uncountable industry:

(1) The clerks in an industry are scattered among a multitude of small, private firms.

(2) The establishment of each workplace in modern industry costs a great deal of capital.

I "error correction"-oppgavene skal studentene lete etter feil og rette dem på skjermen. Her får studenten beskjed om hvor mange feil som eventuelt står igjen når han/hun har sendt inn oppgaven, og lenke til en fasit. Mange av feilene i disse oppgavene er hentet fra autentiske student-essays i NICLE-korpuset.¹

Figur 2 viser et eksempel på en åpen oppgave. Her blir studenten bedt om å resonnere over språkbruk i korpuseksempler og skrive en egen tekst, som siden kan sammenlignes med et "suggested answer". Det er altså en progresjon i oppgavetyperne, fra det relativt mekaniske til mer intellektuelt og akademisk krevende spørsmål.

The screenshot shows a web interface for a concord search task. At the top, there is a navigation bar with a logo 'OIE' and a menu of numbered links (#1 to #8). Below this is a search bar with the text 'Search the corpus:' and a 'Go!' button. The main content area is titled '[Concord #7]' and contains three tasks: Task (1) asks to search for quantifiers 'a series of', 'a group of', and 'a collection of' and determine their concord; Task (2) asks for the motivation for the choice of concord; Task (3) asks to search for 'the series', 'the group', and 'the collection' and determine if they follow the same concord pattern. Below the tasks is a large text input area labeled 'WRITE HERE:'. At the bottom of the interface, there are two buttons: 'Show suggested answer' and 'Undo answer'.

Figur 2: Eksempel på oppgavetyperen "open"

Korpuset

Oppgavene i OIE er som vi har sett korpusdrevne, og tanken er at studentene skal kunne løse alle oppgavene hvis de bruker korpuset på en riktig måte. Noen av oppgavene er lagt opp slik at korpuset *må* brukes for å komme fram til riktig svar, mens det i andre tilfeller kan brukes mer som et oppslagsverk. Studentene skal alltid kunne komme fram til svaret ved å søke i korpuset. Dette medfører at man må ha god kjennskap til og kontroll over tekstene som ligger til grunn for korpuset. Vi valgte derfor å sette sammen et eget korpus istedenfor å bruke et som allerede lå fritt tilgjengelig på nettet. I tillegg utviklet vi også et eget søkegrensesnitt, for at den interaktive dialogen med korpuset skulle bli så enkel som mulig. Uten god kontroll med både innhold og søkegrensesnitt, kan man risikere at søkeresultatene blir endret og ikke stemmer overens med oppgavene.

OIE-korpuset består av ca. 7 millioner ord med engelskspråklige tekster fra det 20. århundre, med noen få unntak der tekstene er av nyere dato. Korpuset inneholder både skjønnlitterære tekster, i form av romaner og noveller, og sakprosa, i form biografier, lærebøker, rapporter, osv. I tillegg kommer taler og filmmanus. Sakprosaen og de skjønnlitterære tekstene er hovedsaklig hentet fra en tidlig versjon av *Longman/Lancaster English Language Corpus* (Summers 1993). Det opprinnelige Longman-korpuset har imidlertid blitt endret noe ved at dupliserte tekster har blitt fjernet, og ved at visse teksttyper har blitt utelatt, som for eksempel matoppskrifter og poesi. Romaner fra 1800-tallet har også blitt tatt ut.

Taler og filmmanus var ikke med i det opprinnelige Longman-korpuset, men ble lagt til for å tilføre korpuset noe (i nærheten av) talespråk. Talene, som alle ble lastet ned fra internett, er i stor grad taler fremført av politikere. Filmmanuskriptene ble også hentet fra ulike nettsteder, og består av britiske og amerikanske filmer.

Noen nye sakprosaetekster er også lagt til; disse er alle rapporter fra ulike lands regjeringer, og ble hentet fra offisielle nettsteder i Australia, Storbritannia og USA. Grunnen til at disse ble inkorporert i korpuset var delvis for å få inn tekster av nyere dato, og delvis for å få en bedre balanse i antall ord mellom sakprosa og skjønnlitteratur. OIE-korpuset er likevel ikke helt balansert når det gjelder fordelingen mellom skjønnlitteratur, sakprosa og talespråk. Dette byr ikke på nevneverdige problemer for OIE, siden oppgavene ikke er lagt opp til at man skal gjøre statistiske sammenligninger på tvers av tekstkategorier. Skal man derimot bruke OIE-korpuset til andre formål der teksttype står sentralt, er det viktig at man er klar over denne forskjellen.

Selv om det meste av materialet i korpuset kommer fra britiske eller amerikanske kilder, er også en del andre engelskspråklige land representert, som for eksempel Sør-Afrika, Australia og Canada.

Korpuset er lagret i en vanlig tekstfil med minimal koding. Hver tekst er utstyrt med en "header", der man kan finne metainformasjon om teksten, f.eks. hvilket land teksten kommer fra og i hvilket år den ble skrevet. Denne relativt enkle formen for koding gjør at korpuset ikke egner seg for spesielt avanserte søk. I neste punkt skal vi se på hvilke søkemuligheter som finnes.

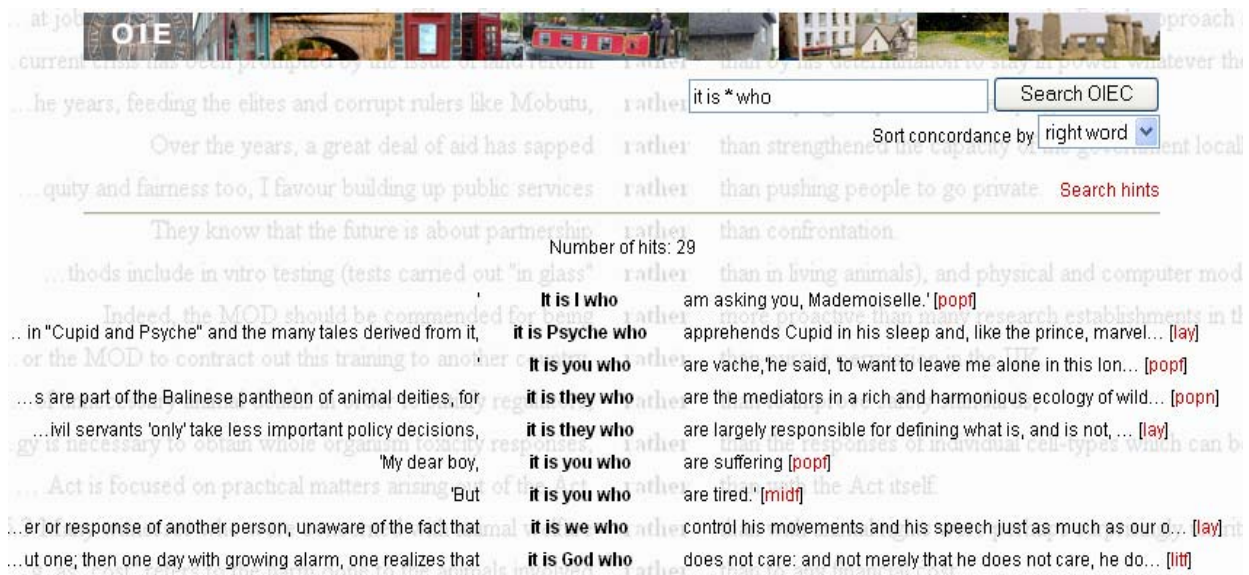
Søking i korpuset

I tillegg til å være et frittstående og fleksibelt tilbud i språklæringen, er OIE på mange måter et innføringskurs i bruk av korpus. Et enkelt søkegrensesnitt med relativt få søkemuligheter ble dermed foretrukket. Studentene har likevel et valg mellom "enkle" og "avanserte" søk. I enkle søk skriver man inn et ord eller en streng av ord i søkevinduet. De litt mer avanserte søkemulighetene tillater søk med et såkalt jokertegn ("wildcard"), dvs. at man setter * på det stedet i søkevinduet man ikke ønsker å spesifisere, f.eks.: [find*]. I tillegg har man også muligheten til å sortere konkordansen på ordet til venstre eller ordet til høyre for søkeordet.

Disse forholdsvis begrensede søkemulighetene gir selvsagt ikke studentene noen innføring i mer kompliserte former for korpussøk, som f.eks. lemmasøk eller søk i predefinerte tekster, men vi tror at et så vidt enkelt og oversiktlig søkegrensesnitt gjør at studentene lettere kan få en følelse av å lykkes. De blir ikke forstyrret av at det er for mange søkekriterier å velge mellom.

Søkeresultatet blir gitt i en KWIC-konkordans, der søkeordet er fremhevet i midten av hver søkelinje. Antall treff i korpuset blir vist foran hver konkordans. På slutten av hver søkelinje finnes en lenke som tar deg til informasjonen som er lagret i headeren for hver fil.

Eksempel 3 viser et søk med jokertegn (it is * who) sortert på ordet til høyre for søkeordet.



The screenshot shows the OIE search interface. At the top, there is a search bar containing the query "it is * who" and a "Search OIEC" button. Below the search bar, there is a dropdown menu for "Sort concordance by" set to "right word". A "Search hints" link is visible. The main area displays a list of search results, each consisting of a snippet of text with the search query highlighted in the center. The number of hits is shown as 29. The results are as follows:

Snippet	Highlighted Query	Position
... am asking you, Mademoiselle.'	it is I who	[popf]
... apprehends Cupid in his sleep and, like the prince, marvel...	it is Psyche who	[lay]
... are vache, he said, 'to want to leave me alone in this lon...	it is you who	[popf]
... are the mediators in a rich and harmonious ecology of wild...	it is they who	[popn]
... are largely responsible for defining what is, and is not, ...	it is they who	[lay]
... are suffering	it is you who	[popf]
... are tired.'	it is you who	[midf]
... control his movements and his speech just as much as our o...	it is we who	[lay]
... does not care: and not merely that he does not care, he do...	it is God who	[litr]

Eksempel 3: KWIC-konkordans i OIE-korpuset

Fra språklæring til språkforskning

Korpusbruk virker spesielt tiltalende i språkundervisning på universitetsnivå fordi skrittet fra undervisning til forskningsmetode er så kort. Studentene vil med andre ord få innsikt i hvordan man kan forske på språk ved hjelp av korpus samtidig som de tilegner seg kunnskaper og ferdigheter i språket. De får dermed et gunstig utgangspunkt for språkstudier på høyere nivåer.

Også en vanlig student, uten forskningsambisjoner, kan benytte de metoder som korpuslingvistikken har utviklet til å gjøre oppdagelser om språket i kontekst. Alle tekstene i korpuset er skrevet av personer som har engelsk som morsmål. Likevel er det viktig at studentene blir gjort oppmerksom på at ikke all språkbruk i korpuset er grammatisk korrekt i følge reglene for standard engelsk. Dette gjelder selvsagt for all bruk av korpus, og det er som alltid viktig at brukerne tolker korpusmaterialet grundig.

Korpuslingvistikken kan på denne måten brukes som en *modell for språklæring*. Ved å gjøre bruk av korpus kan studentene lære å gjøre sine egne oppdagelser om språket, slik at det er studentens egen aktivitet som skaper læringen. Gjennom korpusdata blir studenten eksponert for en mengde autentiske eksempler. Både mengden (kvantiteten) og naturligheten (kvaliteten) av eksemplene muliggjør en effektiv internalisering av språkstrukturer, med rom for flere nyanser og mer variasjon enn en skolegrammatikk kan gi (Mukherjee 2002: 67). Michael Hoey (2004: 37) problematiserer begrepet "authentic data" ved å påpeke at disse vanligvis blir "inauthentically encountered in the classroom, in that they are read or heard for reasons remote from those that gave rise to the data in the first place." Hans konklusjon er likevel – ikke overaskende – at "autentiske data" likevel skal brukes fordi "only authentic data can preserve the collocations, colligations, semantic associations of the language".

Den enkle formen for korpuslingvistikk som OIE representerer, kan samtidig brukes som en *modell for språkforskning*. Den aktive læringsformen som ligger i å trekke slutninger om språkbruk ut fra autentiske eksempler er til forveksling lik det språkforskere gjør når de studerer korpuseksempler. Gjennom OIE-øvelsene blir studenten for eksempel gjort oppmerksom på at relativt synonyme ord (som *big* og *great*) kan ha temmelig ulike kollokasjonsmønstre. Både kollokasjon og "colligation" (det at et ord ofte opptrer sammen med – eller i – en bestemt grammatisk konstruksjon) er sentrale temaer i leksisbasert korpuslingvistikk, for eksempel Sinclair (1991) og Hoey (2004). Det er meningen at OIE skal kunne forberede til forskning ved at progresjonen fra språklæring til språkforskning jo i stor grad ligger i originaliteten og kompleksiteten til problemstillingen snarere enn i valg av metode. Når det gjelder undervisningen i engelsk språk ved Universitetet i Oslo er et også et viktig poeng at mange av lærerne bruker korpuslingvistikk som metode i sin egen forskning. Studenten vil få et innblikk i hvordan en språkforsker arbeider, og kan dermed få følelsen av å bli opplært i "ordentlig" forskning.

På mer avanserte kurs kan studentene få adgang til andre korpus med mer avanserte søkemuligheter for å behandle mer krevende problemstillinger i semesteroppgaver og masteravhandlinger. Universitetet i Oslo har for tiden slike kurs både på bachelor- og på masternivå, hvor man selvsagt er enda nærmere språkforskningen enn man er i OIE-oppgavene. På disse kursene har vi erfaring for at gode studenter ikke bare lærer seg å håndtere korpusmateriale i sine oppgaver, men også bruker korpus som oppslagsverk når de skriver engelsk, og eventuelt senere når de skal undervise i engelsk.

Diskusjon

Fra et pedagogisk synspunkt er det lett å peke på fordeler med korpusbruk i undervisningen. Blant annet blir en student som bruker korpuset i sin språklæring eksponert for rike mengder med autentisk tekst med stor bredde og variasjon i vokabular, grammatiske strukturer og stilnivå (kfr. Mukherjee 2002: 71). Videre gir korpusbruken øvelse i å trekke sine egne slutninger om

språkbruk og vil bidra til å utvikle en generell oppmerksomhet på språklige fenomener. Sist, men ikke minst, er korpusbruk i positiv forstand vanedannende; når man har vent seg til å bruke korpus som hjelpemiddel til å forstå og produsere et fremmedspråk, kan man nesten ikke klare seg uten.

På den annen side kan man også innvende at analyse av korpuseksempler er en krevende øvelse som mest vil komme de beste studentene til gode. Og det er kanskje ikke de studentene som har mest behov for å øve på adjektiv/adverb og samsvar mellom subjekt og verbal. Imidlertid er det tatt høyde for dette ved den progresjonen som er innebygd i OIE-opplegget. Studenter som føler at de ikke trenger de mest elementære øvelsene kan simpelthen hoppe over de første oppgavene.

Når det gjelder de pedagogiske sidene ved e-læring i språkundervisningen generelt, står man overfor en del utfordringer. Den største er fraværet av mellommenneskelig kontakt og bruken av muntlig språk som kommunikasjonsmiddel ved denne typen aktivitet, spesielt når den ikke foregår i et klasserom med lærer/instruktør til stede. Et praktisk problem med integreringen av e-læring i språkundervisningen er at universitetet ikke har mange nok klasserom med pc-plasser til at studentene kan bruke opplegget i timen. Bruken av OIE i undervisningen må derfor begrense seg til at læreren kan demonstrere det og be studentene løse oppgaver til neste time.

Den store utfordringen ved et autonomt nettbasert opplegg som Oslo Interactive English er å få studentene til å ta det i bruk. Foreløpig har opplegget bare vært brukt i et semester (høsten 2005), og erfaringene så langt har vært blandede. Lærerne på det aktuelle nivået har blitt oppfordret til å integrere OIE som del av sin vanlige undervisning, og har fått brukernavn og passord som de kan gi til studentene. Mange studenter har vært innom OIE-sidene, men relativt få har løst særlig mange oppgaver. Det gjenstår altså et arbeid med å motivere studentene til å bruke nok tid på OIE til at de kan dra nytte av korpuset som språklæringsressurs, ikke bare som en "optional extra" som de ikke behøver bruke tid og krefter på.

Avsluttende kommentar

Det er ønskelig at Oslo Interactive English etter hvert blir grundig evaluert, både med tanke på utbyttet studentene får i forhold til forbedret språkferdighet og i hvilken grad de lærer å bruke korpus i språklæring og språkforskning. Hvis resultatet av en slik evaluering er positivt, er det mange måter å bygge ut dette tilbudet på – foruten å fylle på med flere oppgaver og oppgavetyper enn de som allerede ligger der. Det er allerede igangsatt arbeid med et mer spill-preget prosjekt i tilknytning til OIE: *Trivial Corpus Pursuit*. Her er opplegget mye det samme som i OIE, men man samler poeng fra ulike kategorier av spørsmål i en "kake" av samme type som i *Trivial Pursuit*.

Det er klart at opplegg av typen OIE har stor overføringsverdi til undervisning i andre fremmedspråk enn engelsk. Videre kan man også tenke seg en enda enklere variant til bruk i gymnaset – hvis man samtidig lager et opplegg for å lære opp underviserne til å ta det i bruk. Men det er nok en annen historie.

Nettsteder

Oslo Interactive English: www.hf.uio.no/ilos/OIE/ (NB: søking i korpuset krever passord)
Tim Johns Data-Driven Learning page: <http://www.eisu.bham.ac.uk/johnstf/timconc.htm>

Trivial Corpus Pursuit: <http://www.hf.uio.no/ilos/OIE/html/TCPproject.html>

Litteratur

- Hasselgård, Hilde, Stig Johansson og Per Lysvåg 1998. *English Grammar: Theory and Use*. Oslo: Universitetsforlaget.
- Hoey, Michael. 2004. The textual priming of lexis. I Aston, Guy, Silvia Bernardini, Dominic Stewart (red.). *Corpora and Language Learners*, 21-44. Amsterdam: Benjamins.
- Lysvåg, Per. 2005. *The English Language - A Text-based Resource Book: From increased awareness to improved writing*. Kristiansand: Høyskoleforlaget.
- Mukherjee, Joybrato. 2002. *Korpuslinguistik und Englischunterricht. Eine Einführung*. Frankfurt am Main: Peter Lang.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Summers, Della. 1993. Longman/Lancaster English Language Corpus - Criteria and Design. *International Journal of Lexicography* 6 (3): 181-208.

ⁱ NICLE-korpuset er et delkorpus i the International Corpus of Learner English og består av tekster skrevet på engelsk av norske studenter. Se for øvrig <http://www.fltr.ucl.ac.be/FLTR/GERM/ETAN/CECL/Cecl-Projects/Icle/icle.htm>.