

# Instruks for assistenter i TRAWL-prosjektet

Av Hildegunn Dirdal

Sist oppdatert: 20. januar 2023

## Innhold

1. Innsamling av samtykkeskjemaer, spørreskjemaer og tekster .....	1
2. Registrering av metadata .....	2
2.1 Metadata om elevene .....	2
2.2 Metadata om tekstene .....	4
3. Bearbeiding av elektroniske tekster .....	6
3.1 Pdf-kopi av oppgaveteksten .....	6
3.2 Kopi av elevtekstene .....	7
3.3 Ulike utgaver av elevtekstene .....	7
4. Bearbeiding av håndskrevne tekster .....	10
4.1 Pdf-kopi av oppgavetekst .....	11
4.2 Elektronisk versjon av elevtekster .....	11
4.3 Koding av doc-filen .....	13
4.4 Anonymisering av originaltekster .....	14
5. XML-filer med rettet staving .....	15
Tillegg: Spørsmål og svar .....	17

## 1. Innsamling av samtykkeskjemaer, spørreskjemaer og tekster

Samtykkeskjemaer og spørreskjemaer blir distribuert til lærere, elever og foresatte av prosjektmedarbeidere på hvert sted. Hvordan innsamlingen av disse skal foregå, avtales med hver enkelt skole. Hvis de ikke sendes til en av prosjektmedarbeiderne, vil assistentene kunne få i oppgave å hente dem på skolene og følge opp hvis det er noen som mangler. Rutiner avtales med prosjektmedarbeider(e) på stedet for å holde orden på hvilke skoler og klasser de ulike skjemaene er samlet inn fra (ettersom dette ikke står i selve skjemaene).

Hvordan tekstene skal samles inn, vil også bli avtalt med de enkelte skoler/lærere, slik at det blir minst mulig arbeid for lærerne. Hvis det er snakk om håndskrevne tekster, kan det bli assistentenes oppgave å reise ut til skoler et par ganger i semesteret for å kopiere eller skanne disse. Det er da viktig at man noterer/får med:

- 1) hvem som har skrevet hver enkelt tekst,

- 2) når den ble skrevet,
- 3) oppgavetekst/instruks,
- 4) om det er en individuelt skrevet eller samskrevet oppgave,
- 5) om den er skrevet på skolen, hjemme eller som prøve.

Lærerne skal få beskjed om at vi trenger disse opplysningene, men assistentene må sjekke at de foreligger også ved tekster som samles inn på andre måter.

Andre mulige innsamlingsmetoder er at vi får tilgang til innleveringsmapper i ItsLearning eller andre læringsplattformer, at lærere selv kopierer/skanner og sender oss tekster, at lærere videresender tekster elektronisk, osv. Assistentene på hvert sted vil få informasjon om hva som er avtalt med de skolene/lærerne de skal forholde seg til.

I de tilfellene der assistenter har fått tilgang til innleveringsmapper, er det viktig at de laster ned bare tekster fra elever med samtykke.

## 2. Registrering av metadata

Metadata fra spørreskjemaene, samt informasjon om tidspunkt for skriving, versjon av tekst, klassetrinn og oppgavetyper, skal bindes sammen med hver enkelt tekst som samles inn. Assistentenes jobb blir å registrere disse dataene for hver elev og hver tekst som beskrevet nedenfor.

De bruker én Excel-fil for registrering av data fra spørreskjemaet for elevene («Student metadata») og én Excel-fil for registrering av informasjon for tekstene som samles inn («Text metadata»).

### 2.1 Metadata om elevene

Begynn med å legge inn informasjon om elevene fra spørreskjemaene i Excel-filen «Student metadata».

*Name:* Navnet på studenten føres inn først. Dette er informasjon som skal slettes etter at alle tekstene er samlet inn og prosjektet er avsluttet, slik at ingen kan identifisere noen av elevene.

*Pupil code:* Denne koden skal være unik for hver elev slik at man i det ferdige korpuset kan finne ut hvilke tekster som er skrevet av samme elev, men ikke hvem dette var. Vi bruker et femsifret nummer og begynner på et vilkårlig tall for elevene som er med foreløpig, slik at det ikke skal være mulig å vite hvem som var med i begynnerfasen av prosjektet. Assistentene får hvert sitt tall å starte nummereringen fra. Fordi kodeprogrammet ikke kan ta koder som begynner med et tall, settes bokstaven P (for «pupil») først. Eksempel på elevkode: P00347.

*School year/grade:* Før inn klassetrinn som et tosfret tall med «Y» foran, fra «Y05» for femte klasse til «Y13» for siste år på videregående.

*Education program:* Dette er bare aktuelt for elever på videregående. Før inn «N/A» for alle andre. Assistentene holder oversikt over hvilke programmer vi samler tekster fra på co

- Idrettsfag: ID
- Kunst, design og arkitektur: KD
- Medier og kommunikasjon: ME
- Musikk, dans og drama: MD
- Studiespesialisering: ST
- Bygg- og anleggsteknikk: BA
- Design og håndverk / medieproduksjon: DH
- Elektrofag: EL
- Helse- og oppvekstfag: HS
- Naturbruk: NA
- Restaurant- og matfag: RM
- Service og samferdsel: SS
- Teknikk og industriell produksjon: TP

*Level (for foreign languages):* Dette er bare aktuelt for fremmedspråkene (spansk, tysk, fransk). Skriv inn «Level 1», «Level 2» eller «Level 3» alt etter hvilket nivå eleven tar fremmespråk på. Skriv «N/A» hvis du registrerer elever som bidrar med tekster på engelsk.

*Gender:* Før inn «female» hvis det er krysset av for jente/kvinne, «male» hvis det er krysset av for gutt/mann og «other» hvis det er krysset av for annen kjønnsidentitet.

*Birth country, Pupil's L1, Parent/guardian 1: L1, Parent/guardian 2: L1:* Før inn de engelske navnene på landet og språkene som er oppført i spørreskjemaet.

*Have you lived in other countries than Norway?:* Skriv «no» hvis eleven har svart negativt eller ikke har ført opp noen land eller tidspunkter. Skriv «yes» hvis eleven har bodd i andre land.

*If yes, which and when:* Før inn tresifret kode for land etterfulgt av tidsrom i følgende format: «FRA 2014-2015» eller «FRA June 2002-October 2002». Bruk de offisielle kodene du finner på dette nettstedet: <http://www.worldatlas.com/aatlas/ctycodes.htm>. Hvis eleven ikke har bodd i andre land enn Norge, skriver du «N/A» her.

*Languages you can READ/WRITE/SPEAK/COMPREHEND in addition to your L1:* Bruk engelske navn på språkene og list dem opp med komma mellom for hver kategori. Eksempel: English, Arabic, German. Skriv «none» hvis eleven ikke har oppgitt andre språk.

*Have you gone to school with other teaching languages than Norwegian?:* Skriv «no» hvis eleven ikke har oppgitt noen klassetrinn eller undervisningsspråk og «yes» hvis vedkommende har oppgitt klassetrinn og språk.

*If yes, which school year/grade and language?:* Før opp skoleår med et tosifret tall med Y foran og navnet på språk på engelsk. Hvis eleven har oppgitt flere år og språk, skilles disse med komma. Eksempel: «Y03 German, Y06 French». Barnehage føres som «Y00». Hvis

eleven ikke har gått på skole med andre undervisningsspråk enn norsk, fører du opp «N/A» her.

*Which school year did you start with English education?:* Før opp skoleår med et tosifret tall med Y foran.

*Have you lived in an English-speaking country? How long?:* Før opp ett av de følgende ettersom hva som er krysset av: «no, or less than 2 weeks», «2-4 weeks», «1-3 months», «4-6 months», «7-12 months», «more than one year».

*Hours per week spent on ...:* Skriv ett av de følgende alternativene for hver kolonne: «more than 10 hours», «5-10 hours», «1-4 hours», «up to 1 hour», «none». Hvis eleven har satt kryss midt mellom to tidsangivelser, kan du føre dem opp med skråstrek mellom. Eksempel: «5-10 hours/1-4 hours».

*Hours spent on other specified uses of English:* Meld inn slike tilfeller til prosjektmedarbeideren(e) på ditt sted slik at vi kan standardisere valgene. Du vil da få beskjed om hvilke ord/fraser du skal bruke. For lytting til sanger med engelsk tekst skriver du «listen to English songs». Sett inn et komma mellom spesifisering av bruk og angivelse av timeantall (timeantall formuleres som ovenfor). Hvis eleven ikke har ført opp noe her, skriver du «N/A». Hvis eleven har ført opp et timeantall, men ikke spesifisert hva det gjelder, skriver du «unspecified», så f.eks.: «1-4 hours, unspecified».

Dersom informasjon mangler for noen av spørsmålene (ikke bare er uaktuell – N/A) og man ikke kan slutte seg til informasjonen fra andre opplysninger i spørreskjemaet, skriver du «missing» i feltet det gjelder.

## 2.2 Metadata om tekstene

Når tekstene kommer inn, skal de få en tekstkode, og alle metadata om teksten og eleven som skrev den, skal legges inn i Excel-filen «Text metadata». Det er denne filen som vil bli brukt i den siste runden med tekstkoding slik at alle metadataene vil legges inn i en header i tekstfilen og bli søkbare i korpuset. Metadataene om eleven kopieres fra «Student metadata», men husk å oppdatere kolonnene «School year/grade», «Education program» og «Level» dersom dette er endret siden eleven ble med i prosjektet. Dersom vi ikke har spørreskjema fra hvert av årene som eleven har vært med, må man vurdere hvilke opplysninger som er gyldige for de andre årene ut fra når skjemaet er fylt inn. Når det gjelder hvilke språk man kan og hvor ofte man bruker engelsk/fransk/spansk/tysk til ulike aktiviteter, skal det henvises til det året vi har data fra. For eksempel: Hvis eleven har fylt ut spørreskjemaet på åttende trinn og man legger inn tekster fra niende og tiende trinn, skriver man «missing – see Y08».

I tillegg kommer følgende opplysninger:

*Text code:* elevnummer\_klassetrinn\_unikt oppgavenavn\_versjon\_ORIG

Elevnummeret er det samme som «pupil code» beskrevet i 2.1. Klassetrinn oppgis som tosifret tall med Y foran fra «Y05» til «Y13» ettersom vi samler inn tekster fra femte klasse og ut videregående skole. Det unike oppgavenavnet settes basert på oppgaveteksten/innholdet i oppgaven og forkortes til en kode på fire bokstaver (e.g. «POWE» for heldagsprøven med tema «The Power of Language»). Vi må passe på at ulike oppgaver ikke får samme navn selv om de skulle handle om samme/lignende tema, og at samme navn blir brukt hvis nøyaktig samme oppgave er benyttet i ulike klasser. Det er opprettet en Dropbox-side der dere kan legge inn kopi av oppgavetekstene dere får inn og navnet dere har brukt på dem. Hver gang dere får en ny oppgavetekst, må dere gå inn og sjekke om oppgaven finnes fra før og allerede har fått en kode eller om koden dere har tenkt å bruke allerede er brukt. Hvis oppgaven allerede har fått en kode, bruker dere den. Hvis ikke, registrerer dere oppgaven med en ny kode i tabellen i Dropbox. Dokumentet i Dropbox beskriver hvordan dette skal gjøres. Versjon: Bruk «V0» hvis oppgaven bare foreligger i én utgave, «V1» hvis det er et førsteutkast og «V2» hvis det er et andreutkast. Til slutt settes «\_ORIG» for at kodingsprogrammet skal kunne hente ut riktig fil når det legger inn metadata (se kapittel 3 og 4).

Eksempel: P05123\_Y11\_POWE\_V1\_ORIG

*Date submitted:* Bruk følgende format: «D30.09.2016». For håndskrevne tekster fra barneskolen kan det noen ganger være vanskelig å finne dato. Dersom det er umulig å finne datoen for innlevering/skriving av teksten, kan dere bruke ukenummer: «Week 25, 2016». Hvis dette heller ikke foreligger, kan dere skrive måned eller i verste fall bare semester («December 2016» eller «Autumn, 2016»). Vær alltid så spesifikke som dere kan.

*Text language:* Før opp det engelske navnet på språket teksten er skrevet på, e.g. «English», «Spanish», «French», «German» eller «Norwegian» (ettersom vi også kommer til å samle inn norske tekster fra noen elever).

*Text format:* Velg mellom «digital» og «handwritten».

*Text status:* Skriv «V0» hvis oppgaven bare foreligger i én utgave, «V1» hvis det er et førsteutkast og «V2» hvis det er et andreutkast.

*Task type:* Velg mellom «classroom writing», «homework» og «test».

*Teacher comments/feedback:* Skriv «yes» hvis læreren har gitt tillatelse til at kommentarer/tilbakemelding kan lagres og brukes og det faktisk foreligger en kommentert tekst. Skriv «no» hvis læreren ikke har gitt tillatelse til dette eller det ikke foreligger noen kommentarer eller tilbakemelding.

*Teacher code:* Dette er bare aktuelt når det finnes kommentarer som læreren har gitt tillatelse til å bruke, dvs. hvis du har skrevet «yes» i forrige kolonne. Hvis dette ikke er tilfelle, skriver du «N/A» her. Alle lærere som har gitt tillatelse til at kommentarene deres kan brukes, skal tildeles en unik lærerkode. Assistentene får hvert sitt firesifrede tall å starte nummereringen

fra. Fordi kodeprogrammet ikke kan ta koder som begynner med et tall, settes bokstaven T (for «teacher») først. Eksempel på lærerkode: T0056.

### 3. Bearbeiding av elektroniske tekster

Installer VESPA-makroene som beskrevet i «The VESPA Tagging Manual. Version 2.3»

Tekstene vi får inn fra skolene, bør lagres i mapper der det går klart fram at dette er de som kommer direkte fra skolen. Kall gjerne denne mappen «Tekster fra skolene» og lagre dem på minst to lokaliteter. Lag et system av mapper og undermapper, slik at det er enkelt å holde styr på hva som kommer fra elever fra de ulike skolene og klassene.

Gå deretter fram som følger:

#### 3.1 PDF-kopi av oppgaveteksten

Anonymiser oppgaveteksten på følgende måte:

- a. Se om det er navn i teksten som kan identifisere skolen eller lærere. Erstatt i så fall disse med henholdsvis NAME\_PERSON1\_F (for kvinnelige navn), NAME\_PERSON1\_M (for mannlig navn), NAME\_FAMILY1 (for familienavn/etternavn som står alene), NAME\_PLACE1 og NAME\_INSTITUTION1. Hvis det er flere ulike personnavn, fortsetter man med ulike tall etter PERSON e.g. NAME\_PERSON2\_F, NAME\_PERSON3\_M, og tilsvarende for steder eller institusjoner.
- b. Fjern spor etter den som har skrevet dokumentet via programvare i Word: Gå til «File», klikk på knappen «Check for issues» og velg «Inspect document». Du får da spørsmål om du først vil lagre, men kan svare nei på dette. Du får deretter opp en rute der det er haket av for mange ting som dokumentet vil sjekkes for, og du klikker på «Inspect». Hvis det da blir funnet «Document properties and personal information», klikker du på knappen «Remove all» ved siden av dette. Ikke fjern andre ting.

Lag en kopi av oppgaveteksten og lagre den som PDF i en mappe som heter «Oppgavetekst».

Filnavn: klasstrinn\_unikt oppgavenavn.pdf

Eksempel: Y11\_POWE.pdf

Klasstrinn angis med tosifret tall med Y foran fra «Y05» til «Y13». Det unike oppgavenavnet settes basert på oppgaveteksten/innholdet i oppgaven og forkortes til en kode på fire bokstaver (e.g. «POWE» for heldagsprøven med tema «The Power of Language»). (Se informasjon om unikt oppgavenavn i 2.2 ovenfor.)

## 3.2 Kopi av elevtekstene

Lag en kopi av elevtekstene og lagre dem i en mappe som heter «Arbeidstekster». Dette er tekstene du begynner å jobbe med. Gi tekstene følgende filnavn:

Filnavn: elevnummer\_klassetrinn\_unikt oppgavenavn\_versjon\_ORIG

Eksempel: P05123\_Y11\_POWE\_V1\_ORIG

Filnavnet skal være det samme som tekstkoden (text code) for oppgaven (se 2.2).

## 3.3 Ulike utgaver av elevtekstene

For hver elevtekst skal det lages flere utgaver på følgende måte:

### 3.3.1 Åpne elevteksten og anonymiser den ved å:

- a. Fjerne navn, skole o.l. i topptekst
- b. Se om det er navn i teksten som kan gjøre at eleven identifiseres, e.g. navn på personer, dyr, steder eller institusjoner som kan knyttes til eleven. Erstatt i så fall disse med henholdsvis NAME\_PERSON1\_F (for kvinnelige navn), NAME\_PERSON1\_M (for mannlig navn), NAME\_FAMILY1 (for familienavn/etternavn som står alene), NAME\_ANIMAL1\_F (for feminint dyrenavn), NAME\_ANIMAL1\_M (for maskulint dyrenavn), NAME\_PLACE1 og NAME\_INSTITUTION1. Hvis det er flere ulike personnavn, fortsetter man med ulike tall etter PERSON e.g. NAME\_PERSON2\_F, NAME\_PERSON3\_M, og tilsvarende for ulike dyr, steder eller institusjoner. Da er det lettere å holde styr på referentene i den anonymiserte teksten. (Bruk ulike nummer selv om kjønnen er forskjellig.) Det kan også forekomme telefonnumre eller fødselsdatoer som må erstattes med henholdsvis NUMBER\_PHONE og NUMBER\_DATE (med tall etter hvis det er flere ulike som for navn ovenfor). Hvis læreren har lagt inn kommentarer i dokumentet, må disse også sjekkes for personlige opplysninger. Noen lærere nevner f.eks. navnet til studenten. Hvis du er sikker på at et navn ikke har noe med eleven å gjøre og ikke kan brukes til å identifisere ham/henne, lar du det stå. Hvis du er i tvil, bytter du ut navnet som forklart ovenfor.
- c. Fjerne spor etter den som har skrevet dokumentet via programvare i Word: Gå til «File», klikk på knappen «Check for issues» og velg «Inspect document». Du får da spørsmål om du først vil lagre, men kan svare nei på dette. Du får deretter opp en rute der det er haket av for mange ting som dokumentet vil sjekkes for, og du klikker på «Inspect». Hvis det da blir funnet «Document properties and personal information», klikker du på knappen «Remove all» ved siden av dette. Ikke fjern andre ting.

### 3.3.2 Klikk på «save» for å lagre dokumentet i anonymisert form, men hold det åpent.

3.3.3 Hvis det er kommentarer fra læreren i dokumentet, og hvis læreren har gitt tillatelse til at disse kan brukes, skal du nå lagre en PDF-versjon der disse inngår. Lagre teksten som PDF-fil i en mappe som heter «Anonymiserte originaltekster med kommentarer».

Filnavn: elevnummer\_klassetrinn\_unikt oppgavenavn\_versjon\_COMM.pdf

Eksempel: P05123\_Y11\_POWE\_V1\_COMM.pdf

Filnavnet skal være det samme som tekstkoden (text code) for oppgaven (se 2.2) bortsett fra at «\_ORIG» byttes ut med «\_COMM».

3.3.4 Lukk PDF-filen, men jobb videre med teksten du har åpen. Fjern eventuelle endringer gjort av lærer ved å gå til «Review», trykke på pilen under tasten med «Reject» og velge «Reject All Changes in Document». Deretter fjerner du alle kommentarer ved å trykke på pilen under tasten med «Delete» og velge «Delete All Comments in Document». Endringer og kommentarer kan også fjernes i Document Inspector. Men vær obs på at hvis læreren har skrevet inn rettelser i selve teksten til elevene i Google Docs, blir lærerens rettelse da stående, mens elevens opprinnelige tekst, det som er streket gjennom, forsvinner. I disse tilfellene må man fjerne lærerkommentarene og -endringene via «Review». Dette har vist seg å gjelde tekster som er lastet ned fra Classrooms, og man bør sjekke nøye hva som er tilfelle dersom man får filer som er lastet ned fra nye plattformer. Til slutt lagrer du den anonymiserte teksten uten lærerkommentarer i en mappe som heter «Anonymiserte originaltekster».

Filnavn: elevnummer\_klassetrinn\_unikt oppgavenavn\_versjon\_ORIG.pdf

Eksempel: P05123\_Y11\_POWE\_V1\_ORIG.pdf

Filnavnet skal være det samme som tekstkoden (text code) for oppgaven (se 2.2), bortsett fra at det er en PDF-fil.

3.3.5 Lukk PDF-filen, og lagre deretter teksten som doc-fil i en mappe som heter «Tekster for koding». Filnavnet skal ellers være det samme.

Filnavn: elevnummer\_klassetrinn\_unikt oppgavenavn\_versjon\_ORIG.doc

Eksempel: P05123\_Y11\_POWE\_V1\_ORIG.doc

(Grunnen til at vi bruker \_ORIG (for «original») selv om teksten nå er anonymisert og kommentarer tatt bort, er at vi senere skal lage korrigerede versjoner av tekstene, der staving blir rettet. Elevenes originale tekster vil derfor bli lenket til korrigerede tekster og stå i kontrast til disse.)

3.3.6 Teksten som nå er en doc-fil skal du kode ved å bruke makroene.

Makroene er opprinnelig utviklet for et korpus som heter VESPA, og brukes til å legge inn koder for kursiv og fet skrift, overskrifter, setninger, avsnitt, kapitler, kildelister, sitater, nevnte ord og fraser, bilder/tegninger/figurer, tabeller og lister.

Det er en separat manual for bruken av VESPA-makroene. Det er viktig at du leser denne nøye!

Velg Add-Ins på verktøylinja. Da får du fram to knapper du kan trykke på: én for VESPA: man (manuell koding) og én for VESPA: auto (automatisk koding). Start med manuell koding slik det ble gjennomgått på opplæringsmøtet og slik manualen beskriver det. Kjør til slutt



automatisk koding. Når du har klikket OK på beskjeden «Auto-tagging finished for the file», er det kodede dokumentet automatisk lagret i mappen «tagged2» som en XML-fil.

### **Merk følgende når dere koder tekstene:**

Vi kaller alle overskrifter på essay-nivå «Section 1» uavhengig av om det er ett eller flere essay i filen, og bruker bare «document title» hvis det er en overskrift av typen «eksamensoppgave i ...» med flere deler/essay. Hvis det er underoverskrifter innenfor samme essay, bruker vi «Section 2» på disse. Hvis et essay/en oppgave ikke har noen overskrift, bruker vi bare koden «start of running text».

Fordi taggen for sitater (<q>...</q>) ikke kan legges inn i en setnings-tag, bruker vi denne bare for frittstående sitater (block quotes). Alle andre sitater kodes som «mentioned items» Denne taggen kan stå inni en setning.

Tekst sitert fra et annet sted er det eneste som skal kodes som mentioned items, for eksempel: Harmer (2016: 21) writes: «Motivation can be defined as ...», ... when he sings that «the flower was crushed»... Hvis du ser at eleven siterer, skal du kode sitatet som mentioned item selv om hermetegn mangler eller om eleven har brukt kursiv eller annet til å vise at noe er et sitat. Både titler på og sitater fra sanger, bøker, o.l. skal kodes som mentioned items. Følgende skal IKKE kodes som mentioned items: tilfeller der eleven bruker hermetegn for å vise at han/hun bruker et ord med forbehold (scare quotes), replikker i en historie eleven har diktet opp, ord/fraser som er gitt i oppgaven eleven svarer på.

NB: Makroene er satt opp slik at de forstår alt fra der man markerer fram til første “paragraph mark” som overskrift, uansett om det er en “section” eller “bibliography”. Det er derfor viktig at man sjekker at det er en “paragraph mark” der overskriften slutter. Den skal se slik ut:



Hvis elevene har brukt “manual line break” i stedet, blir ting feil. “Manual line break” ser slik ut:



Når man starter makroene, slår maskinen automatisk på visning av slike tegn, som dere sikkert har lagt merke til. Hvis dere ser at det er brukt “manual line breaks”, må disse byttes ut med “paragraph mark” etter overskrifter eller ved slutten av avsnitt. De må byttes ut med mellomrom hvis de ser ut vil å være brukt i midten av en setning el.l. bare for å få startet på en ny linje (dette skjer nok ikke så ofte). Se tillegget med spørsmål og svar nederst i dokumentet for mer informasjon om hvordan dette kan gjøres og om andre detaljer ved kodingen.

“Page breaks”, “section breaks” og Word styles (brukt f.eks. til å markere overskrifter) må også fjernes (som det står på s. 6 i “The VESPA tagging manual”).

Siden det ikke er noen kode for emojis/emoticons i taggeren, skal disse byttes ut med ordet EMOJI. Dette må gjøres manuelt. Skriv EMOJI én gang for hver emoji/emoticon som er brukt.

Eventuelle kommentarer til de som skal bruke korpuset, skal skrives inn i det lille vinduet for «comments» som åpnes automatisk når man begynner tagging-prosessen (se 2.3.9 i VESPA-manualen). Alle kommentarer skal skrives på engelsk. Meld inn forhold som går igjen og som dere synes det er naturlig å legge inn kommentarer på, så kan vi lage standard-formuleringer. Foreløpig har vi følgende standard-formuleringer:

- The tagged text is part of a larger set of tasks. Translation tasks and listening- or reading-comprehension tasks have not been tagged, but are available in the anonymized original text document.
- The original document contained notes to the answers, which have not been included in the tagged text, but are available in the anonymized PDF file.

Når du er ferdige med et oppgavesett fra en klasse, laster du opp alle filene (PDF-filer, doc-filer, tagged og tagged2) i prosjektets lagringsområde eller overfører dem til prosjektlederen ettersom hva som er avtalt ved din institusjon.

Husk: Ikke slett arbeidstekstene før prosjektet er ferdig. De er viktige å ha som back-up.

(Tekstene vil bli behandlet videre med et Perl-skript utviklet for VESPA og tilpasset til vårt korpus. I dette trinnet legges det inn flere koder, alle kodene konverteres til såkalte TEI-koder, og Perl-skriptet henter metadata fra Excel-filen «Text metadata» og legger dette inn i en header i den endelige XML-versjonen. Dette er også beskrevet i VESPA-manualen.)

## 4. Bearbeiding av håndskrevne tekster

På lavere nivåer skriver elevene ofte tekster for hånd. Noen lærere vil selv kopiere disse og sende dem til oss; i andre tilfeller reiser assistentene ut til skolen for å gjøre denne jobben.

For at kopieringen skal gå greit og effektivt samtidig som man holder orden på tekstene, er det lurt å forberede noen enkle mapper for hver elev som har sagt ja til å bidra med tekster.

Mappene får du av din prosjektleder. Skriv på navn og hvilken klasse elevene tilhører. Gjør klare noen ark for hver klasse der du kan noterer hvilke oppgaver de har svart på, når tekstene ble skrevet/levert, om de ble skrevet individuelt eller i gruppe og hva slags oppgave det var (hjemmelekse, skrevet på skolen eller prøve). Ta med de samme arkene neste gang du drar til skolen slik at du vet hva som var siste tekst som ble kopiert.

Ta kopi av oppgaveteksten hvis denne er fra læreboka eller ikke er kopiert opp av læreren.

Kopier alle elevtekstene som har blitt skrevet siden sist du var på skolen. Det må være minst én setning for at det skal regnes som en tekst. Vi samler altså ikke inn innfyllingsoppgaver og lignende, og heller ikke oversettelsesoppgaver. Du får med deg kopipapir av din prosjektleder, slik at du ikke behøver å bruke skolens papir til dette.

Før du begynner å bearbeide tekstene, må du lage et nytt sett med kopier av dem alle. Lever dette settet til prosjektmedarbeideren på ditt sted, og bruk det settet du beholder når du skriver inn teksten elektronisk og lager en anonymisert utgave som forklart nedenfor.

Lag et system av elektroniske mapper og undermapper til å lagre elektroniske utgaver av tekstene i, slik at det er enkelt å holde styr på hva som kommer fra elever fra de ulike skolene og klassene.

#### 4.1 PDF-kopi av oppgavetekst

Skann oppgaveteksten (eller skriv den inn i en Word fil – se nedenfor) og lagre den som PDF i mappen for gjeldende oppgave i en undermappe som heter «Oppgavetekst».

Filnavn: klasstrinn\_unikt oppgavenavn.pdf

Eksempel: Y05\_FORE.pdf

Klasstrinn angis med tosifret tall med Y foran fra «Y05» til «Y13». Det unike oppgavenavnet settes basert på oppgaveteksten/innholdet i oppgaven og forkortes til en kode på fire bokstaver (e.g. «FORE» for oppgaven «A trip to the forest»). (Se informasjon om unikt oppgavenavn i 2.2 ovenfor.)

Hvis oppgaven går ut på å lese en tekst og skrive noen setninger om den, skal dere ta en kopi av teksten også. Vi skal finne ut hva vi gjør i forhold til copyright etter hvert, men da har vi i alle fall en kopi. Skriv ned referanse til boka den er tatt fra. Når det gjelder oppgaveordlyd som kommer fra læreboka, kan dere skrive den inn i et Word-dokument. Noter om det hørte til noe bilde eller lignende, og skriv inn full referanse til boka med sidetall. (Følg Chicago author–date-systemet.)

Eksempel:

Thorsen, Cecilie and Hege Dahl Unnerud. 2008. *Stairs 7: Textbook*. Oslo: Cappelen Damm. (Page 25)

Hvis det ikke finnes noen oppgavetekst eller du ikke finner den, lager du et unikt oppgavenavn basert på innholdet i teksten og registrer dette i Dropbox-filen på vanlig måte. Lag et dokument der du skriver inn «Assignment text missing» og lagre dette dokumentet som du ville lagret oppgaveteksten.

#### 4.2 Elektronisk versjon av elevtekster

Skriv inn hver elevtekst elektronisk. Teksten skal skrives akkurat slik som eleven har gjort det, selv om den skulle inneholde stavfeil, tegnsettingsfeil eller grammatikkfeil. Husk å slå av automatiske retteprogrammer slik at ikke Word retter på ting som skal være feil. Hvis det er steder der du ikke kan tolke hva eleven har skrevet setter du ordet/ordene du tror det kan være i klammeparentes. Dette gjelder også om det bare er en av bokstavene i ordet du ikke kan gjenkjenne. Hvis det er umulig å se hvilket ord det er, setter du et spørsmålstegn i parentes.

Eksempel 1: Det kan være vanskelig å se forskjell på «u» og «a» hos en elev, og du lurer på om han har skrevet «cup» eller har feilstavet ordet som «cap». Da transkriberer du som følger: He was playing in the [cup].

Eksempel 2: Det er et ord du ikke klarer å tolke. Da skriver du en klammeparentes med spørsmålsteget i: We played [?] all morning.

Hvis det er navn i teksten som kan gjøre at eleven kan identifiseres, e.g. navn på personer, dyr, steder eller institusjoner som kan knyttes til eleven, skal disse ikke skrives inn i den elektroniske versjonen, men erstattes med henholdsvis NAME\_PERSON1\_F (for kvinnelige navn), NAME\_PERSON1\_M (for mannlig navn), NAME\_FAMILY1 (for familienavn/etternavn som står alene), NAME\_ANIMAL1\_F (for feminint dyrenavn), NAME\_ANIMAL1\_M (for maskulint dyrenavn), NAME\_PLACE1 og NAME\_INSTITUTION1. Hvis det er flere ulike personnavn, fortsetter man med ulike tall etter PERSON e.g. NAME\_PERSON2\_F, NAME\_PERSON3\_M, og tilsvarende for ulike dyr, steder eller institusjoner. Da er det lettere å holde styr på referentene i den anonymiserte teksten. (Bruk ulike nummer selv om kjønnet er forskjellig.) Det kan også forekomme telefonnumre eller fødselsdatoer som må erstattes med henholdsvis NUMBER\_PHONE og NUMBER\_DATE (med tall etter hvis det er flere ulike som for navn ovenfor). Noen ganger kan angivelser av aldre muligens være identifiserende, for eksempel hvis hele familien er listet opp med alder. Da kan koden NUMBER\_AGE brukes. Hvis du er sikker på at et navn ikke har noe med eleven å gjøre og ikke kan brukes til å identifisere ham/henne, lar du det stå. Hvis du er i tvil, bytter du ut navnet som forklart ovenfor.

Hvis det er tegninger inne i teksten, skriver du DRAWING på det gjeldende stedet. Hver tegning markeres for seg der den står, men hvis det er tegninger rundt teksten, f.eks. i marginen, markeres de bare som én tegning (uansett om de er relatert til teksten eller ikke).

Emojis/emoticons erstattes med EMOJI (én gang for hver emoji/emoticon som er brukt).

Hvis det er to deler som svarer på samme oppgave, skriver du disse inn i samme dokument.

Pass på at dere bruker «paragraph mark» ved linjeskift, ikke «manual line break» for at kodingen med makroene skal bli riktig. «Page breaks», «section breaks» og Word styles (brukt f.eks. til å markere overskrifter) må heller ikke brukes. Se «Merk følgende når dere koder tekstene» under punkt 3.3.6 ovenfor samt spørsmål og svar i tillegget nederst i dokumentet.

Lagre teksten som en doc-fil i en mappe som heter «Tekster for koding». Gi filen navn etter følgende mønster:

Filnavn: elevnummer\_klassetrinn\_unikt oppgavenavn\_versjon\_ORIG.doc

Eksempel: P05600\_Y05\_FORE\_V0\_ORIG.doc

Filnavnet skal være det samme som tekstkoden (text code) for oppgaven (se 2.2).

### 4.3 Koding av doc-filen

Teksten som nå er en doc-fil skal du kode ved å bruke makroene. Makroene er opprinnelig utviklet for et korpus som heter VESPA, og brukes til å legge inn koder for kursiv og fet skrift, overskrifter, setninger, avsnitt, kapitler, kildelister, sitater, nevnte ord og fraser, bilder/tegninger/figurer, tabeller og lister.

Det er en separat manual for bruken av VESPA-makroene. Det er viktig at du leser denne nøye!

Velg Add-Ins på verktøylinja. Da får du fram to knapper du kan trykke på: én for VESPA: man (manuell koding) og én for VESPA: auto (automatisk koding). Start med manuell koding slik det ble gjennomgått på opplæringsmøtet og slik manualen beskriver det.

Hvis det var tegninger i teksten som du har skrevet inn som «DRAWING», krysser du av for at du vil kode Tables, Figures and Graphs når du får opp den første boksen. Når du kommer til valgene som gjelder slike ting, koder du «DRAWING» som Picture.

Hvis det var to tekster som svarte på samme oppgave, kodes overskriften på hver av dem som «section 1».

Fordi taggen for sitater (<q>...</q>) ikke kan legges inn i en setnings-tag, bruker vi denne bare for frittstående sitater (block quotes). Alle andre sitater kodes som «mentioned items» Denne taggen kan stå inni en setning.

Tekst sitert fra et annet sted er det eneste som skal kodes som mentioned items, for eksempel: Harmer (2016: 21) writes: «Motivation can be defined as ...», ... when he sings that «the flower was crushed»... Hvis du ser at eleven siterer, skal du kode sitatet som mentioned item selv om hermetegn mangler eller om eleven har brukt kursiv eller annet til å vise at noe er et sitat. Både titler på og sitater fra sanger, bøker, o.l. skal kodes som mentioned items. Følgende skal IKKE kodes som mentioned items: tilfeller der eleven bruker hermetegn for å vise at han/hun bruker et ord med forbehold (scare quotes), replikker i en historie eleven har diktet opp, ord/fraser som er gitt i oppgaven eleven svarer på.

Eventuelle kommentarer til de som skal bruke korpuset, skal skrives inn i det lille vinduet for «comments» som åpnes automatisk når man begynner tagge-prosessen (se 2.3.9 i VESPA-manualen). Alle kommentarer skal skrives på engelsk. Meld inn forhold som går igjen og som dere synes det er naturlig å legge inn kommentarer på, så kan vi lage standard-formuleringer. Foreløpig har vi følgende standard-formuleringer:

- The tagged text is part of a larger set of tasks. Translation tasks and listening- or reading-comprehension tasks have not been tagged, but are available in the anonymized original text document.
- The original document contained notes to the answers, which have not been included in the tagged text, but are available in the anonymized PDF file.

Kjør til slutt automatisk koding. Når du har klikket OK på beskjeden «Auto-tagging finished for the file», er det kodede dokumentet automatisk lagret i mappen «tagged2» som en XML-fil.

#### 4.4 Anonymisering av originaltekster

Gå tilbake til papirkopien du har av den håndskrevne teksten og anonymiser den. De navnene som kunne identifisere eleven, og som du erstattet med «NAME\_PERSON1\_M, NAME\_PLACE1», osv. når du skrev inn teksten elektronisk, skal dekkes over slik at de ikke er lesbare (ved hjelp av korrekturlakk, svart tusj, eller annen tildekking). Dette gjelder også navn i eventuelle kommentarer fra læreren.

4.4.1 Hvis det er kommentarer fra læreren i dokumentet, og hvis læreren har gitt tillatelse til at disse kan brukes, skal du nå lagre en PDF-versjon der disse inngår. Skann teksten og lagre den som PDF-fil i en mappe som heter «Anonymiserte originaltekster med kommentarer».

Filnavn: elevnummer\_klassetrinn\_unikt oppgavenavn\_versjon\_COMM.pdf

Eksempel: P05600\_Y05\_FORE\_V0\_COMM.pdf

Filnavnet skal være det samme som tekstkoden (text code) for oppgaven (se 2.2) bortsett fra at «\_ORIG» byttes ut med «\_COMM».

4.4.2 Ta deretter igjen for deg papirkopien og dekk over alle kommentarer fra læreren slik at de ikke er lesbare (ved hjelp av korrekturlakk, svart tusj, eller annen tildekking). Nå skanner du den anonymiserte teksten uten lærerkommentarer og lagrer den i en mappe som heter «Anonymiserte originaltekster».

Filnavn: elevnummer\_klassetrinn\_unikt oppgavenavn\_versjon\_ORIG.pdf

Eksempel: P05600\_Y05\_FORE\_V0\_ORIG.pdf

Filnavnet skal være det samme som tekstkoden (text code) for oppgaven (se 2.2).

Når du er ferdige med et oppgavesett fra en klasse, laster du opp alle filene (PDF-filer, doc-filer, tagged og tagged2) i prosjektets lagringsområde eller overfører dem til prosjektlederen ettersom hva som er avtalt ved din institusjon.

Husk: Ikke slett arbeidstekstene før prosjektet er ferdig. De er viktige å ha som back-up.

(Tekstene vil bli behandlet videre med et Perl-skript utviklet for VESPA og tilpasset til vårt korpus. I dette trinnet legges det inn flere koder, alle kodene konverteres til såkalte TEI-koder, og Perl-skriptet henter metadata fra Excel-filen «Text metadata» og legger dette inn i en header i den endelige XML-versjonen. Dette er også beskrevet i VESPA-manualen.)

## 5. XML-filer med rettet staving

Ut fra hva som er avtalt ved din institusjon, vil du enten få tilsendt XML-filene ('tagged3' og tagged3 CORR') eller laster dem ned fra prosjektets Teams-rom. I 'tagged3' ligger ORIG-filene, og i 'tagged3 CORR' ligger filene hvor du skal rette skrivefeil. Du skal ikke røre ORIG-filene unntatt hvis instruksene (under) sier noe annet.

Åpne filene med Oxygen XML Editor. Det lønner seg å åpne både ORIG-filen og CORR-filen samtidig, disse vil da vises i en fane over vinduet der du kan redigere teksten. På denne måten kan du gjøre endringer parallelt hvis det trengs. Hvis man har stor nok PC-skjerm, er det mulig å åpne begge filene ved siden av hverandre i Oxygen med synkronisert scrolling, noe som vil gjøre rettingsarbeidet lettere. Gå da til "Window" og velg "Tile editors vertically", og velg "synchronous scrolling".

Det første du gjør er å klikke på ikonet som heter 'Format and indent' på oppgavelinjen over tekstvinduet for å slå på tekstbryting, slik at du kan se all teksten.

En liten firkant vil vises oppe i høyre hjørne av tekstvinduet. Hvis den er grønn, ble filen kjørt gjennom Perl uten feil. Hvis den er rød, er det en feil i filen, og det må finnes ut av.

Hvis alt er i orden, kan du nå begynne å rette teksten. Scroll ned forbi all metadata-info til du kommer til elevens tekst, som starter med <body>.

Det som er viktig å forstå når man nå skal begynne å rette, er at det må være like mange ord i begge filene, ORIG og CORR. Hvis det at vi retter et ord i CORR da medfører at det blir flere eller færre ord her enn i ORIG, må vi da passe på at vi kompenserer ved å legge til et ekstra ord i en av filene. Se instruksjonene under.

### **Engelsk – instruksjoner for retting:**

- Skrivefeil skal rettes. Vi godtar variasjon i eksisterende stavemåter, f.eks. når det er forskjell på amerikansk/britisk staving av samme ord. Bruk OneLook.com for å kjøre et søk på ordet på tvers av online ordbøker hvis du er usikker. Det er også mulig å åpne teksten i Word for å få opp en versjon med stavekontroll. I tilfeller hvor du er i tvil om det er stavefeil eller feil ordvalg, retter du ikke feilen. Bruk sunn fornuft.
- Feil bruk av stor/liten bokstav skal rettes.
- Grammatikkfeil og tegnsettingsfeil skal IKKE rettes, med følgende unntak:
- Hvis et aksent-tegn ( ` , ´ ) er brukt i stedet for en apostrof ( ´ ) skal dette rettes i *både* ORIG- og CORR-filene. Man bør bruke søk/erstatt-funksjonen i Oxygen til å finne og erstatte de stedene der aksent-tegn er brukt i stedet for apostrof, slik at dette ikke overses.

- Hvis det mangler en apostrof i et ord hvor det skal være apostrof, skal denne fylles inn i CORR-filen, for eksempel før genitivs-s (*John's*) eller i sammentrekninger (*don't*). Legg da til [[CON]] etter ordet i ORIG-filen. Dette er fordi den automatiske taggeren som skal analysere ordklasser, deler opp ord der det er apostrof. Hvis man setter inn en apostrof i CORR-filen, vil taggeren se ordet som to ord. Derfor må det inn en [[CON]] i ORIG-filen slik at det blir like mange ord der.
- Hvis det er satt inn en apostrof der den ikke skal være, for eksempel før flertalls-s, skal denne tas vekk i CORR-filen. Du må da kompensere med å legge til [[SPL]] etter ordet i CORR-filen slik at antall ord ikke forandrer seg ift. ORIG-filen. Dette trenger du ikke å gjøre dersom apostrofen er plassert etter ordet, slik som i *Dickens'*.
- Dersom eleven feilaktig har skrevet to ord sammenslått (f.eks. *parkbench* i stedet for *park bench*), legg til [[CON]] etter ordet i ORIG-filen og splitt ordene i CORR-filen. Siden vi splitter opp et ord til to i CORR-filen, trenger vi altså et ekstra ord i ORIG-filen slik at det blir like mange ord i begge, derfor setter vi inn [[CON]] i ORIG-filen.
- Dersom eleven feilaktig har delt opp et ord som skulle vært sammenslått (f.eks. *basket ball* i stedet for *basketball*), legg til [[SPL]] etter ordet i CORR-filen og slå sammen ordet i CORR-filen. Denne sammenslåingen i CORR-filen medfører at det blir ett ord mindre der enn i ORIG, derfor må da vi sette inn [[SPL]] i CORR-filen.
- Feil bruk av bindestrek skal rettes. Ord som er bundet sammen med bindestrek regnes som ett ord, så dersom er brukt bindestrek der det ikke skal være en bindestrek (f. Eks. *South-Africa*), må det settes inn [[CON]] i ORIG-filen når dette rettes. Dersom det mangler bindestrek der en bindestrek skal være, må du da sette inn [[SPL]] i CORR-filen når dette rettes,
- Dersom du oppdager feil i VESPA-taggingen, retter du disse ved å skrive inn de riktige taggene i både CORR- og ORIG-filen. De riktige taggene kan du kopiere fra et annet sted i teksten, eventuelt en annen XML-fil.
- Husk at hvis du bruker søk- og erstatt-funksjonen til å rette skrivefeil som dukker opp ofte, så vær forsiktig slik at du ikke endrer på ord i kodingen eller metadataen i teksten (slik som for eksempel hvis du endrer *form* til *from*).

Når du er ferdig med å rette, lagrer du filen. Når alle filene i mappene er ferdig rettet, lastes mappene opp igjen i oppgavemappen i Teams (erstatt da mappene som lå der fra før).



## Tillegg: Spørsmål og svar

### Om tagging:

#### Makroene ser ut til å tolke et helt avsnitt som en overskrift – hvorfor skjer dette?

Svar:

Makroene er satt opp slik at de forstår alt fra der man markerer fram til første “paragraph mark” som overskrift, uansett om det er en “section” eller “bibliography”. Det er derfor viktig at man sjekker at det er en “paragraph mark” der overskriften slutter. Den skal se slik ut:

¶

Hvis elevene har brukt “manual line break” i stedet, blir ting feil. “Manual line break” ser slik ut:

↵

Når man starter makroene, slår maskinen automatisk på visning av slike tegn, som dere sikkert har lagt merke til. Hvis dere ser at der er brukt “manual line breaks”, må disse byttes ut med “paragraph mark” etter overskrifter eller ved slutten av avsnitt. De må byttes ut med mellomrom hvis de ser ut vil å være brukt i midten av en setning el.l. bare for å få startet på en ny linje (dette skjer nok ikke så ofte).

Her ser dere et eksempel der en elev har brukt “manual line break” etter overskriften, og at makroen har tolket hele det første avsnittet som overskriften på section 1 selv om jeg bare markerte “Task 1 “Why learn English?””:

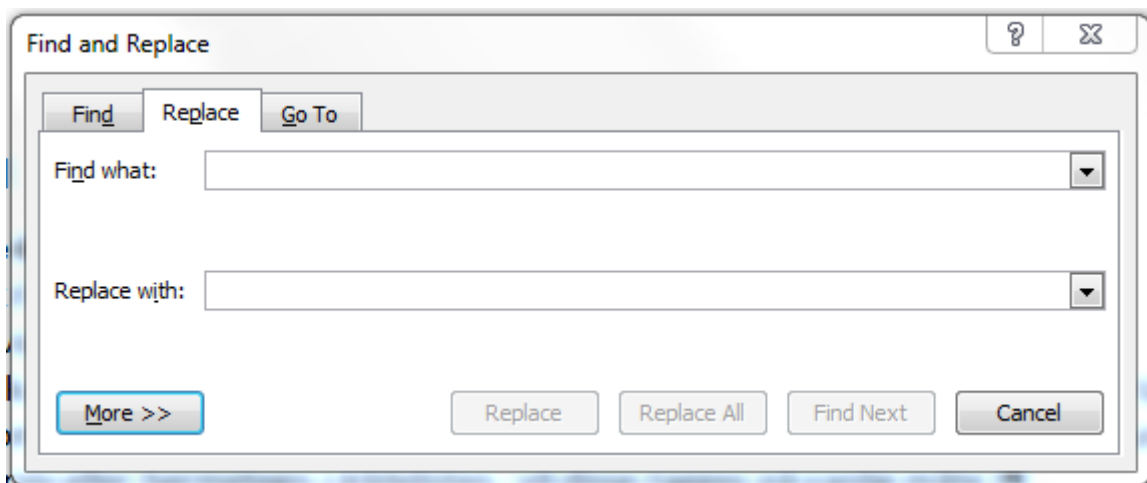
```
{start:vespa_section-1}¶  
Task 1↵  
“Why learn English?”↵  
↵  
The English language is an official language in 52 countries and there are about 350 million people that have English as their mother tongue. It has been considered as Lingua Franca, which basically means a language used for communication between two people with different native language. In some countries, people that don't even have English as their native language use English for business, tourism and other stuff. English is the third most common language in the world. There are various reasons why we should learn to speak English. But let's talk about how English became the global language and why it is so important to understand and speak English.¶  
{end:vespa_section-1}¶  
¶  
English was spread fast and it was adopted by many different countries. From the time of the reign of Queen Elizabeth I and the following 400 years, England power grew and so did the
```

Når “manual line breaks” er byttet ut med “paragraph marks”, blir det riktig:

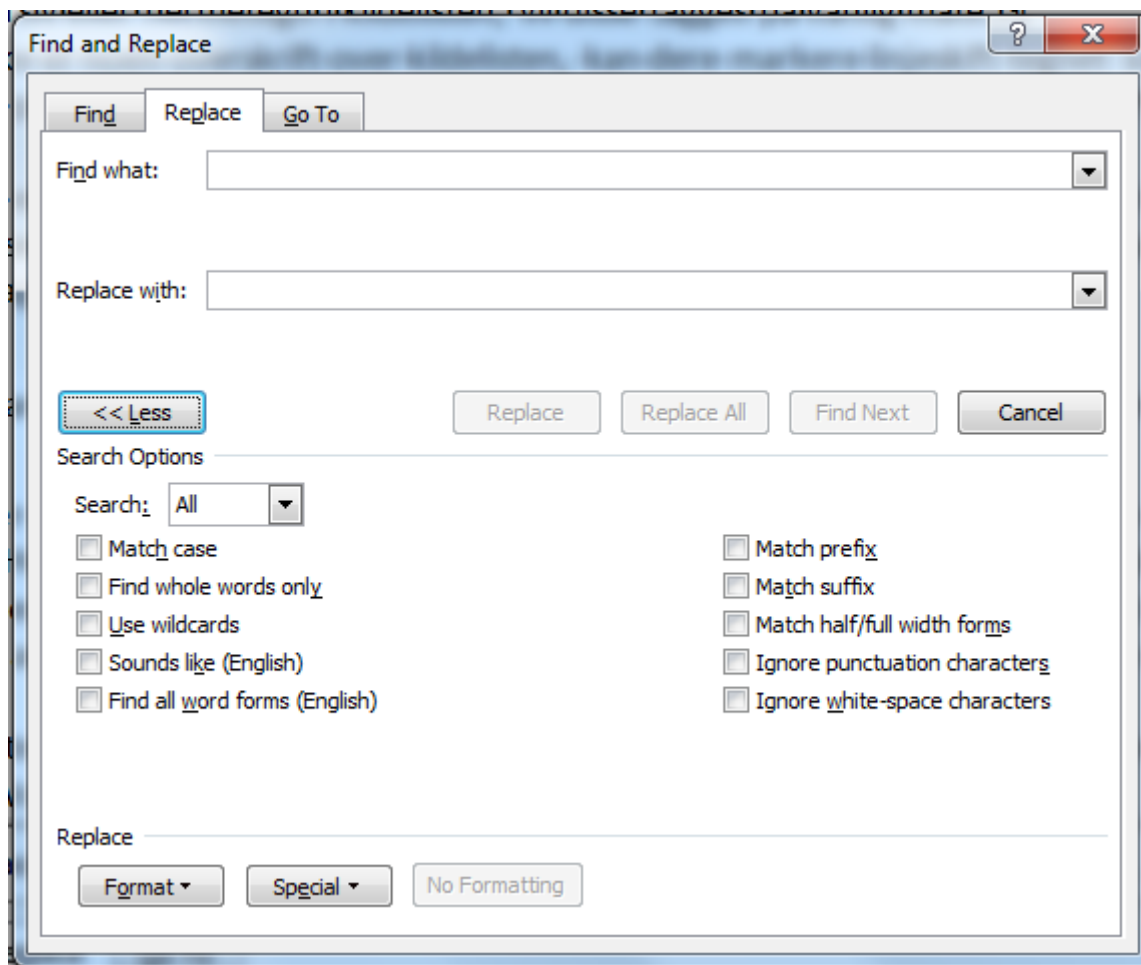
```
{start:vespa_section-1}¶  
Task 1¶  
“Why learn English?”¶  
{end:vespa_section-1}¶  
¶  
The English language is an official language in 52 countries and there are about 350 million  
people that have English as their mother tongue. It has been considered as Lingua Franca,  
which basically means a language used for communication between two people with different  
native language. In some countries, people that don't even have English as their native  
language use English for business, tourism and other stuff. English is the third most common  
language in the world. There are various reasons why we should learn to speak English. But  
let's talk about how English became the global language and why it is so important to  
understand and speak English.¶  
¶  
English was spread fast and it was adopted by many different countries. From the time of the  
reign of Queen Elizabeth I and the following 400 years, England power grew and so did the
```

Når dere har tagget en overskrift som “Bibliography”, “Section 1”, osv. skal altså start-taggen stå rett før overskriften og end-taggen rett etter – ikke lenger nede i dokumentet.

For å bytte ut “manual line breaks” med “paragraph marks” mer effektivt, kan man bruke “replace”, som ligger helt øverst til høyre på verktøylinja under “Home” i Word. Klikk på “Replace” i verktøylinja, og du får frem følgende:



Klikk på “More” nederst, for å få flere valg:



Plasser skrivemerket/kursoren i feltet for “Find what” og klikk på “Special” nederst. Da får du fram en liste der du velger “Manual line break”.

Hvis du skal bytte ut med “paragraph mark” plasserer du skrivemerket i feltet for “Replace with”, trykker på “Special” igjen, og velger “Paragraph mark”.

Hvis du skal bytte ut med et mellomrom, trykker du bare på mellomromstasten etter at du har plassert skrivemerket i “Replace with”-feltet.

Hvis alle “manual line breaks” skal byttes ut med det samme tegnet, kan man trykke på “Replace all”. Hvis ikke, må man ta én og én med “Find next” og “Replace”.

Dette er litt plundrete, men er veldig viktig for at makroene skal gjøre ting riktig. “Page breaks”, “section breaks” og Word styles (brukt f.eks. til å markere overskrifter) må også fjernes (som det står på s. 6 i “The VESPA tagging manual”).

**Hvordan skal vi tagge kilder i kildelista? Overskriften skal tagges som bibliography, men hva med selve kildene? Skal de tagges som mentioned items? Hva om det ikke er noen overskrift over kildelista?**

Svar:

Selve kildene skal ikke tagges. Som med andre “sections”, er det bare overskriften som skal tagges, og Perl-skriptet vi kjører til sist, skjønner da at det som følger etter overskriften hører inn under denne. (Hvis det er kursiv eller hermetegn i kildelisten, vil disse tagges på vanlig måte.)

Dersom det ikke er noen overskrift over kildelisten, kan dere markere “paragraph mark” som står der som overskrift på kildelisten. Programmet vil da forstå at det som følger er en kildeliste.

Det blir seende slik ut:

```
{start:vespa_biblio-1}¶  
¶  
{end:vespa_biblio-1}¶  
http://essaynparagraph.com/importance-of-learning-english/¶  
http://toeflgencligi.blogspot.no/2008/11/importance-of-english.html¶  
http://www.globaltalent.ca/w\_step1/englishskills/¶  
http://english-tonight.com/why-do-people-learn-english-9-most-common-reasons/¶  
http://www.usingenglish.com/articles/importance-english-language.html¶
```

**Noen elever har skrevet kilder i header eller footer. Kan jeg la dem stå der? Eller skal jeg kopiere dem over til enden av dokumentet?**

Svar:

Header og footer vil forsvinne i prosessen, så hvis kildene står der, må de kopieres over til slutten av dokumentet.

**Hvordan skal vi dele inn tekster i sections når de består av svar på et sett med spørsmål og er inndelt etter 1,2,3 og a,b,c? Og hva skal vi gjøre hvis vi vet at oppgaveteksten har en slik inndeling, men eleven ikke har markert ulike deler av svaret?**

Svar:

Poenget med taggingen er at formateringen som en elev har brukt, skal være “synlig” i XML-filen som lages til slutt. Det er altså ikke strukturen på spørsmålet, men hvordan eleven har valgt å strukturere/formatere svaret sitt som gjelder. Hvis eleven ikke har noen inndelinger eller overskrifter, blir det altså ingen koding i “sections” selv om spørsmålet skulle ha flere adskilte deler.

Hvis derimot svarteksten har inndelinger med a., b., c. osv., er det naturlig å tolke disse som en type overskrifter. Men: Man er nødt til å sette inn et “paragraph mark” etter “a.” osv. for at dette skal kunne markeres som en overskrift i seg selv. Hvis ikke, tar makroen med alt fram til første “paragraph mark” den finner, og da blir teksten i oppgaven plutselig definert som overskrift. Dessuten blir det problemer med overlappende sections (en ny section begynner før den første er slutt, osv.) Disse måtene å tagge på ville blitt avvist av Perl-skriptet og vi måtte ha begynt å søke etter feilene.

En eksempeltekst med slike inndelinger og en tagget versjon av denne teksten er lagt ut i Dropbox. Legg spesielt merke til innsatte “paragraph marks” (dere må trykke på tasten for dette i verktøylinja for å se dem hvis dere ikke har den på allerede). Hvis studentene har bruk automatisk nummerering som i eksempelteksten, kan det hende man ikke bare kan trykke på enter for å sette inn paragraph mark, men at man må sette inn tallene/bokstavene på nytt for å få det riktig.

**Skal bilder slettes når de tagges som bilde? Bilder kan limes inn på forskjellige måter, og når de slettes kan det bli store tomrom i dokumentet. Skal vi da fjerne tomrommet?**

Svar:

Bilder skal markeres og tagges som bilde. De vil da automatisk slettes i løpet av taggeprosessen og erstattes med en tag. Se s. 19 i “The VESPA tagging manual”.

**Er det viktig å beholde formatet? Kan vi klippe ut teksten og lime den inn i et nytt dokument dersom eleven har skrevet i tekstbokser eller spalter eller har skrevet i topp- og bunntekst?**

Svar:

Topp- og bunntekst (header og footer) vil forsvinne, så hvis noe som tilhører oppgaven er skrevet der (ikke bare navn, dato, osv.), må det kopieres over.

Hvis det som står i en tekstboks er del av selve besvarelsen og ikke bare en slags illustrasjon, må det også kopieres over.

“Section breaks” må tas vekk, som nevnt ovenfor. Da vil spalter forsvinne.

En kopi av originalen vil jo lagres separat. Der kan man se hvordan det så ut opprinnelig.

## **Mentioned items**

Noen ganger skjer det at sitater får feil tagger. I eksempel nedenfor er det et sitat som har fått feil tag. Det er to «end»-tagger der i stedet for én «start»-tag og én «end»-tag. Pass på at dere

retter opp slike feil, for hvis ikke vil det komme feilmeldinger når vi kjører Perl-skriptet. Pass også på at hermetegnene kommer innenfor taggene (de første står utenfor her, antagelig fordi det er blitt en «end»-tag).

**g.** The American he met at MacDonnell's advises him: "{/q}I was resisting India.  
Once I gave in, I did much better."{/q}

### **Om studentmetadata:**

**Hvis en elev har glemt å fylle inn informasjon som vi med sikkerhet kan slutte oss til fra andre steder (f.eks. kjønn), gjør vi det?**

Svar:

Ja, dere kan fylle inn informasjon dere er sikre på. Informasjon dere ikke kan finne svar på fra andre deler av spørreskjemaet, fører dere som «missing».