



Kartlegging av behovet for automatisk oversettelse i statlig sektor

Utarbeidet for Kommunal- og moderniseringsdepartementet

OE-rapport 2016-15

Om Oslo Economics

Oslo Economics utreder økonomiske problemstillinger og gir råd til bedrifter, myndigheter og organisasjoner. Våre analyser kan være et beslutningsgrunnlag for myndighetene, et informasjonsgrunnlag i rettslige prosesser, eller et grunnlag for interesseorganisasjoner som ønsker å påvirke sine rammebetingelser. Vi forstår problemstillingene som oppstår i skjæringspunktet mellom marked og politikk.

Oslo Economics er et samfunnsøkonomisk rådgivningsmiljø med erfarne konsulenter med bakgrunn fra offentlig forvaltning og ulike forsknings- og analysemiljøer. Vi tilbyr innsikt og analyse basert på bransjeerfaring, sterk fagkompetanse og et omfattende nettverk av samarbeidspartnere.

Om Amesto

Amesto Translations er Norges største og en av Skandinavias ledende leverandører av oversettelsestjenester. Amesto har kontorer i Norge, Sverige og Danmark og hadde en omsetning i 2014 på 41,3 MNOK i Norge og på 113 MNOK totalt i Skandinavia.

I Norge har Amesto 24 ansatte innenfor Translations og våre medarbeidere består av engasjerte og dyktige prosjektledere, statsautoriserte translatører, fagoversettere, faglitterære oversettere og tolker. Totalt har Amesto Translations 70 ansatte i Skandinavia.

Kartlegging av behovet for automatisk oversettelse i statlig sektor/2016-15

© Oslo Economics, 29. april 2016

Kontaktperson:

Asbjørn Englund / Partner

aen@osloeconomics.no, Tel. 913 18 802

Innhold

Sammendrag og konklusjoner	4
1. Om oppdraget	6
1.1 Metode og informasjonskilder	6
2. Beskrivelse av bruksområder og målgrupper for MT@EC	8
2.1 Hva er MT@EC?	8
2.2 Bakgrunn	8
2.3 Hva er maskinoversettelse?	8
2.4 Beskrivelse av MT@EC	9
2.5 Dagens anvendelser av MT@EC	9
2.6 Fremtidige anvendelser av MT@EC i CEF.AT	11
3. Vurdering av behov for MT@EC i statlige virksomheter	12
3.1 Statlige virksomheters oversettelsesbehov	12
3.2 Ressursbruk knyttet til oversettelse i statlige virksomheter	15
3.3 Bruk av digitale hjelpemidler til oversettelse	17
3.4 Behov for oversettelse til/fra norsk i EUs webportaler	18
3.5 Behov for oversettelse til/fra nynorsk i MT@EC	21
3.6 Samlet vurdering av behovet for MT@EC	22
4. Alternativer for en nasjonal satsing på norsk i MT@EC	25
4.1 Språk og fagområder	25
4.2 Satsingsalternativene	26
5. Kost-/nytteanalyse av ulike satsingsnivå på MT@EC	30
5.1 Identifiserte nytte- og kostnadsvirkninger av en satsing på MT@EC	30
5.2 Sammenhengen mellom corpora-størrelse og nytteverdi	33
5.3 Vurdering av ikke-prissatte virkninger	35
5.4 Vurdering av prissatte virkninger ved en satsing på MT@EC	38
5.5 Usikkerhet	40
5.6 Samlet vurdering og anbefaling	41
6. Vedlegg	42

Sammendrag og konklusjoner

For å sikre et fungerende digitalt indre marked med mer informasjonsdeling og kommunikasjon på tvers av landegrensler, har EU-kommisjonen utviklet et verktøy for maskinoversettelse, MT@EC. Systemet er kostnadsfritt for offentlig administrasjon i medlemsland, inkludert Norge, og deres offentlige virksomheter (inkludert deres ansatte). Statlige virksomheter, samt kommuner og deres underliggende virksomheter har dermed fri tilgang til verktøyet.

Formålet med denne utredningen er å gi KMD et beslutningsgrunnlag for å avgjøre nivået på en nasjonal satsing på EU-kommisjonens verktøy for maskinoversettelse, MT@EC. Vi har i utredningen vurdert behovet for maskinoversettelse i statlige virksomheter og gjennomført en kost/nytteanalyse knyttet til innføring av MT@EC.

MT@EC fungerer ved at systemet lærer hvordan enkeltord og setninger skal oversettes basert på innmatede samlinger av oversatte tekster (corpora). Kvaliteten på MT@ECs maskinoversettelse er bedre, jo større corpora er. Foreløpig har norsk bokmål blitt introdusert i systemet basert på den foreliggende oversettelsen av samtlige lovttekster i EØS-avtalen. Ettersom størrelsen på eksisterende corpora er relativt begrenset, er kvaliteten på oversettelsene til/fra norsk bokmål ved hjelp av MT@EC i dag relativt dårlig. For å få systemet til å fungere optimalt vil det påløpe en investeringskostnad knyttet til å skaffe til veie eller produsere corpora. En viktig del av utredningen har derfor vært å avgjøre i hvilken grad man skal prioritere å øke corpora på bokmål fra dagens nivå. I tillegg har vi vurdert hvorvidt man bør utvide verktøyets anvendelsesområde til å også kunne oversette til/fra nynorsk.

Vi har i utredningen innhentet informasjon og innspill fra et stort utvalg departement og etater, samt interne oversettere i statlige virksomheter gjennom intervjuer og spørreundersøkelse.

Behovet for oversettelse i departementer, etater og øvrige virksomheter er varierende ettersom de opererer innenfor et bredt spekter av virksomhetsområder og fagområder. I mange virksomheter er imidlertid behovet stort og flere virksomheter har interne oversettere. Dagens ressursbruk til oversettelse av tekst mellom norsk bokmål og EU-språkene er en viktig indikator på behovet for å ta i bruk maskinoversettelse. Vi anslår at statlige virksomheters ressursbruk knyttet til oversettelse av tekst mellom EU-språk er i størrelsesorden 40-50 millioner kroner årlig. Dette inkluderer kostnader til innkjøp av oversettelse og ressursbruk til interne oversettere (sysselsetter totalt 32 årsverk). I tillegg kommer ressursbruk for oversettelser som utføres av vanlige ansatte. Våre undersøkelser viser at bruk av gratis oversettelsesverktøy som Google Translate og Bing Translate er relativt utbredt blant ansatte i statlige virksomheter.

Mange virksomheter er positive til å ta i bruk maskinoversettelse. Behovet for oversettelser er i mange tilfeller større enn det man oversetter i dag. Det tyder på at det er behov for å effektivisere prosessen med å oversette tekster. Maskinoversettelse kan også være et nyttig hjelpemiddel for å dele informasjon og dokumenter med andre land. Blant de interne oversetterne vurderer vi at behovet for MT@EC er noe begrenset. Bakgrunnen er at de fleste allerede har tatt i bruk profesjonelle oversettelsesverktøy. Vi vurderer derfor at MT@EC vil gi begrensede gevinster utover det de allerede har oppnådd ved å ta i bruk slike verktøy.

I den videre utviklingen av MT@EC vil EU-kommisjonen fokusere på å benytte verktøyet til å gjøre digitale tjenester og plattformer, som støttes eller administreres av EU, multispråklige. MT@EC er allerede integrert i fem slike digitale tjenester/plattformer, men det oversettes foreløpig ikke til/fra norsk. Det foreligger konkrete planer om integrering av maskinoversettelse i ytterligere seks tjenester/plattformer. Eksempelvis lanseres det i nær fremtid en online plattform for forbrukerklager (ODR) som gjør at norske forbrukere kan klage på kjøp av en vare eller tjeneste online fra en næringsdrivende i et annet land. Maskinoversettelse gjør at norsk forbrukere kan skrive inn og motta all informasjon på norsk bokmål. En satsing på bokmål i MT@EC vil bidra til å øke tilgjengeligheten til disse tjenestene ved at brukere kan benytte tjenestene på norsk bokmål. En høyere kvalitet på maskinoversettelsen vil dessuten øke kvaliteten på disse tjenestene/plattformene.

En satsing på MT@EC vil ikke være et «enten eller»-valg, men heller en vurdering av i hvilken grad man burde satse. For å tydeliggjøre mulighetsrommet for en eventuell norsk satsing på MT@EC, har vi delt analysen inn i tre ulike satsingsalternativer. I satsingsnivå 1 baserer man seg på en innsamling av eksisterende corpora i rett teknisk filformat fra oversettere i både offentlige og private virksomheter. I satsingsnivå 2, vil det i tillegg innhentes oversatte tekster også i andre tekniske filformat enn de som kan benyttes direkte av MT@EC (worddokumenter, etc). Dette er mer kostbart ettersom det krever at tekstene bearbeides. Satsingsnivå 3 innebærer at det i tillegg

produseres oversettelser på utvalgte fagområder. Dette gjør at man kan utvide anvendelsesområdene til MT@EC. Vi har også vurdert en satsing på oversettelse til nynorsk i MT@EC.

Den mest vesentlige gevinsten av en satsing på MT@EC er mer kommunikasjon og informasjonsdeling på tvers av landegrenser. Vi forventer at et økt satsingsnivå vil bidra til større gevinster på dette området ettersom den generelle bruken av verktøyet forventes å øke og ved at man kan anvende MT@EC på stadig flere områder. Standardisering av språk og mer konsistent språkbruk er andre positive virkninger man kan oppnå ved å satse på MT@EC.

MT@EC kan imidlertid også ha negative virkninger i form av lavere språklig kvalitet. Maskinoversettelse kan i verste fall bidra til flere misforståelser og mer støy i kommunikasjonen på tvers av landegrenser dersom tekst oversettes feil, det ikke benyttes riktig fagterminologi eller dersom det mangler viktige ord i maskinoversettelsen. Maskinoversettelse bør derfor ikke benyttes ukritisk.

Vi estimerer at kostnaden for en minimumssatsing på MT@EC vil være 3 millioner kroner i netto nåverdi over en 10-årsperiode. Vi vurderer at de positive effektene MT@EC medfører i form av økt informasjonsutveksling på tvers av landegrenser kan forsvare en satsing utover dagens nivå.

En satsing utover satsingsnivå 1 vil gi en betydelig relativ kostnadsøkning, samtidig som økningen i omfanget av corpora er liten. Vi vurderer på bakgrunn av dette at satsingsnivå 1 fremstår som mest hensiktsmessig. Et vesentlig poeng her er at satsingsnivå 1 ikke utelukker muligheten for en videre satsing i fremtiden. Dermed kan satsingsnivå 2 og 3 anses som opsjoner. Med satsingsnivå 1 vil man trolig få en kvalitetsforbedring i oversettelsene på en rekke fagområder. Etter en vurdering av gevinstrealiseringen som følge av satsingsnivå 1, vil det være mulig å investere ytterligere i tekstinnsamling og produksjon av corpora ved behov. Ved produksjon av corpora kan man da prioritere de områdene hvor behovet for kvalitetsforbedringer er stort.

Vi har vurdert en satsing på nynorsk uavhengig av avgjørelsen om bokmålsatsingen. I dag finnes det i liten grad egnet corpora på nynorsk. En nynorsksatsing vil kreve at et stort volum tekst oversettes for å få et tilstrekkelig stort corpus. Det gjør at en satsing på nynorsk blir relativt kostbar. Sammenliknet med det anbefalte alternativet i bokmålsatsingen, satsingsnivå 1, er den estimerte kostnaden til nynorsksatsingen nesten fire ganger så stor. Basert på svarene til virksomhetene og de ansatte i spørreundersøkelsene tillegges oversettelser mellom EU-språk og nynorsk liten verdi. En høy kostnad, kombinert med et lavt rapportert behov for nynorskoversettelser gjør at det fremstår som lite hensiktsmessig å investere i produksjonen av corpora på nynorsk. En slik satsing vil likevel kunne være forsvarlig dersom muligheten til nynorskoversettelser tillegges svært høy egenverdi. Samlet sett vurderer vi at det ikke er grunnlag for å anbefale en satsing på nynorsk i MT@EC.

1. Om oppdraget

Oslo Economics og Amesto Translations har fått i oppdrag av Kommunal- og moderniseringsdepartementet (KMD) å kartlegge behovet for automatisk oversettelse i statlige virksomheter. Formålet med utredningen er å gi KMD et bedre beslutningsgrunnlag for å avgjøre nivået på en nasjonal satsing på et verktøy for automatisk oversettelse utviklet i EU-kommisjonen, MT@EC.

Mandatet for oppdraget har vært å:

- Fremskaffe en overordnet oversikt over ressursbruk knyttet til oversettelse til og fra offisielle EU-språk i statlige virksomheter.
- Kartlegge behovet i statlige virksomheter for MT@EC, og vurdere hvilke gevinster en videre satsing på MT@EC vil kunne gi.
- Vurdere kostnader knyttet til forbedring av bokmål/nynorsk oversettelse i MT@EC.
- Produsere en kost/nytteanalyse knyttet til innføring av MT@EC i norske statlige virksomheter.

Utredningsoppdraget er gjennomført i perioden desember 2015 til og med april 2016.

1.1 Metode og informasjonskilder

Kost/nytteanalysen er gjennomført i henhold til Direktoratet for økonomistyring (DFØ) sin veileder i samfunnsøkonomisk analyse¹. I arbeidet med utredningen har vi benyttet flere metoder for informasjonsinnhenting. I det følgende gir vi en oversikt over metodene vi har benyttet.

1.1.1 Dokumentstudier

Vi har innhentet og gjennomgått informasjon om MT@EC og maskinoversettelse generelt. Dette omfatter blant annet:

- Presentasjon til Spyridon Pilos (Madrid 26.01.2016)
- Brochure MT@EC (European Commission v/DG Translate& DG Informatics)
- Presentasjon EC: «Machine translation for public administrations in the EU member states en»

Videre har vi innhentet informasjon om andre oversettelsesverktøy², for å sikre en grundig forståelse av automatisk oversettelse. Vi har også benyttet EU-

kommisjonens nettsider, samt informasjon om deres webportaler fra nettsidene.

1.1.2 Dybde- og gruppeintervjuer

Utvalget vi har intervjuet inkluderer interne oversettere i statlige virksomheter, en forsker på språkteknologi, ansatte med kjennskap til EUs webportaler og sentrale personer i EU-kommisjonen.

Vi har gjennomført dybdeintervjuer med totalt 4 personer i EU-kommisjonen, 19 personer fra 13 ulike statlige departement og etater, samt forsker Trond Trosterud ved Universitetet i Tromsø.

Formålet med dybdeintervjuene har vært å kartlegge arbeidsmetodene til de profesjonelle oversetterne, deriblant bruken av oversettelsesverktøy. Videre har vi ønsket å undersøke behovet for forbedringer i slike verktøy, hvilke gevinster en satsing på MT@EC kan gi, samt hvilken motivasjon oversetterne har for å ta verktøyet i bruk.

Gjennom intervjuene har vi også søkt å kartlegge i hvilket omfang corpora finnes i virksomhetene. I kartleggingen undersøkte vi blant annet dokumentenes tekniske format, juridiske spørsmål knyttet til avlevering og estimerer av kostnader knyttet til overlevering av corpora og bruken av eksterne oversettere. Følgende norske virksomheter var representert i dybdeintervjuene:

1. Utenriksdepartementet (UD):
 - Avdeling for EØS-oversettelser
 - Seksjon for oversettelser
 - Europaavdelingen
2. Statens vegvesen
3. Forbrukerrådet
4. Mattilsynet
5. Statistisk sentralbyrå (SSB)
6. Sjøfartsdirektoratet
7. Justisdepartementet (JD)
8. Forbruker Europa
9. Barne- og likestillingsdepartementet (BLD)
10. Domstolsadministrasjonen
11. Stortinget

Intervjuene med nøkkelpersoner i EU-kommisjonen er benyttet til å opparbeide en grunnleggende forståelse av MT@EC. For å opparbeide en forståelse av tekniske aspekter ved verktøyet intervjuet vi prosjektets tekniske leder, samt personen som er hovedansvarlig for teknisk utvikling av systemet.

¹

http://www.dfo.no/Documents/FOA/publikasjoner/veileder_e/Veileder_i_samfunns%C3%B8konomiske_analyser_1409.pdf

² Eksempelvis har vi benyttet dokumentet: Benitez, M.T. (2008) *Open architecture for multilingual parallel texts*.

Deretter innhentet vi informasjon om praktisk bruk og arbeid med verktøyet fra den svenske oversettelsesavdelingen i EU-kommisjonen DG Translate. I tillegg kartla vi videre økonomiske og politiske planer knyttet til MT@EC gjennom et intervju med leder for «Research and Innovation sector» i DG Connect. Dermed har vi søkt å bekrefte informasjon fra flere ulike kilder og synspunkter.

Språkforsker Trond Trosterud ved Universitetet i Tromsø ble intervjuet for å belyse MT@EC fra et forskningsperspektiv.

1.1.3 Spørreundersøkelser

Spørreundersøkelse blant statlige virksomheter

Vi har også sendt ut et spørreskjema til en rekke etater og departementer for å kartlegge behovet for MT@EC hos et bredere utvalg av etater og departementer enn i dybdeintervjuene.

Undersøkelsen ble sendt ut til 36 virksomheter, og vi mottok svar fra følgende 29:

1. Arbeids- og sosialdepartementet (ASD)
2. Brønnøysundregistrene
3. Datatilsynet
4. Direktoratet for økonomistyring (DFØ)
5. Direktoratet for byggkvalitet (DIBK)
6. Direktoratet for samfunnssikkerhet og beredskap (DSB)
7. Stiftelsen Miljømerking (Ecolabel)
8. Finansdepartementet (FIN)
9. Finanstilsynet
10. Forbrukerombudet
11. Havarikommisjonen
12. Helse- og omsorgsdepartementet (HOD)
13. Integrerings- og Mangfoldsdirektoratet (IMDi)
14. Innovasjon Norge
15. Klima- og miljødepartementet (KLD)
16. Kripos
17. Kunnskapsdepartementet (KD)
18. Lovdata
19. Miljødirektoratet
20. NAV
21. Nasjonal kommunikasjonsmyndighet (NKOM)
22. NOKUT
23. Nærings- og fiskeridepartementet
24. Olje- og energidepartementet (OED)
25. Petroleumstilsynet (Ptil)
26. Riksrevisjonen
27. Samferdselsdepartementet (SD)
28. Skattedirektoratet
29. Utlendingsdirektoratet (UDI)

Spørreundersøkelsen inneholdt 10 spørsmål (15 inkludert delspørsmål) og omhandlet de samme

temaene som dybdeintervjuene, nemlig bruken av profesjonelle oversettere, ressursbruk knyttet til oversettelser til/fra EU-språk, behov for MT@EC, hvilke språk virksomheten har behov for oversettelse til, mulige gevinster ved å ta i bruk MT@EC og bruken av EU-webportaler og behovet for oversettelse i disse.

Undersøkelse blant ansatte i statlige virksomheter

I tillegg har vi gjennomført en spørreundersøkelse, som har rettet seg mot en bredere gruppe ansatte i statlige virksomheter. Målgruppen for denne undersøkelsen har vært ansatte som ikke er profesjonelle oversettere.

Undersøkelsen ble sendt ut til 102 personer i 8 forskjellige departement og etater, se Tabell 1-1. Vi mottok 45 svar, noe som medfører en svarprosent på 44 prosent. Erfaringsmessig er svarprosenten på denne typen undersøkelser relativt lav, tilsvarende som vi opplevde med vår undersøkelse. Ettersom svarprosenten og utvalget er begrenset, anser vi ikke resultatene som representative for alle ansatte i offentlige virksomheter. Likevel mener vi at undersøkelsen har verdi som kvalitative innspill i våre analyser.

Tabell 1-1: Utvalg spørreundersøkelse blant ansatte

Virksomhet	Antall svar	Antall utsendt
Mattilsynet	19	22
BLD	6	17
Innovasjon Norge	3	18
Forbruker Europa	2	3
Statens vegvesen	2	25
Sjøfartsdirektoratet	4	4
Finanstilsynet	3	4
UDI	6	9
Totalt	45	102

Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

Spørreundersøkelsen inneholdt 47 spørsmål og søkte å kartlegge behovene til ansatte i offentlige virksomheter og «skjulte» kostnader knyttet til oversettelser. Slike kostnader kan eksempelvis bestå av tid brukt på å oversette enkeltdokumenter, der det ikke benyttes profesjonelle oversettere. Videre ønsket vi å kartlegge kjennskap til og bruken av ulike webportaler knyttet til EU. Undersøkelsen omhandlet også behovet for et maskinoversettelsesverktøy, hvilke språk det oversettes til/fra og mulige gevinster ved å ta i bruk MT@EC.

2. Beskrivelse av bruksområder og målgrupper for MT@EC

2.1 Hva er MT@EC?

I EU i dag finnes 24 offisielle språk og 3 alfabeter. For å sikre et fungerende digitalt indre marked, med mer informasjonsdeling og kommunikasjon på tvers av landegrensene og språk har EU-kommisjonen utviklet et verktøy for maskinoversettelse, MT@EC.

MT@EC er en statistikkbasert form for maskinoversettelse som baserer seg på åpen-kilde programvaren Moses. Moses er i dag verdens mest brukte programvare for statistikkbasert maskinoversettelse. MT@EC oversetter i dag mellom 24 offisielle EU-språk. MT@EC baserer seg på et minne bestående av 900 millioner setninger som primært er produsert av oversettere i EU-institusjoner de siste to tiårene.

MT@EC har vært i drift siden juni 2013 og benyttes i dag av oversetteravdelingene i EU og øvrige EU-institusjoner, så vel som offentlig administrasjon i EU/EØS-land. MT@EC er også tilgjengelig i enkelte europeiske digitale tjenester/plattformer som er finansiert eller støttet av EU-kommisjonen. I 2014 ble ca. 7 millioner sider med tekst oversatt ved hjelp av MT@EC.³

MT@EC er kostnadsfritt for medlemsland, inkludert Norge, og deres offentlige virksomheter. I intervjuer har EU-kommisjonen opplyst at de antar MT@EC vil være kostnadsfritt så lenge «Connecting Europe Facility (CEF)»-programmet eksisterer. Automatisk oversettelse er foreløpig sikret finansiering gjennom CEF på 15 millioner euro i perioden 2016-2019. I intervjuer har EU-kommisjonen opplyst om at det i fremtiden kan bli aktuelt å lansere en brukerbetalt løsning for private aktører. Dette er imidlertid usikkert.

EU-kommisjonen arbeider nå med å utvide anvendelsesområdene til MT@EC og videreføringen av MT@EC går under navnet «CEF Automated Translation» (CEF.AT).

EU-kommisjonen vurderer selv at de forventede gevinstene ved MT@EC er:

- Tilrettelegging for enklere informasjonsdeling mellom parter (rettslige samarbeidspartnere, etc.)
- Rask forståelse for innholdet i inngående skriftlig informasjon på andre EU-språk slik at man kan videreformidle den til rett person eller avdeling uten å måtte bestille en komplett skriftlig oversettelse
- Reduserte kostnader siden profesjonelle oversettere kun trenger å oversette de deler av innkommende dokument som anses som viktig.

Kilde: EU-kommisjonen⁴

2.2 Bakgrunn

MT@EC ble utviklet i perioden 2010 til 2013 av oversetterseksjonen i EU-kommisjonen. Bakgrunnen for at MT@EC ble utviklet var at EU-kommisjonen erfarte at man på tross av betydelig ressursinnsats på oversettelsesområdet (i 2013 var det 1700 oversettere i EU-kommisjonen som oversatte 2 millioner sider årlig) fortsatt hadde et mye større behov for oversettelser enn det oversetterne klarte å dekke. Eksempelvis ville en oversettelse av europa.eu til alle offisielle EU-språk kreve sysselsetting av 8500 oversettere per år og oversettelse av 6,8 millioner dokumenter. For å sikre at enda flere dokumenter blir oversatt har derfor EU-kommisjonen valgt å utvikle et verktøy for maskinoversettelse.

2.3 Hva er maskinoversettelse?

Det har de siste årene skjedd store framskritt når det gjelder anvendelse av teknologi for å oversette tekst. Ulike dataprogrammer som fasiliteter oversetterens arbeid har blitt stadig mer utbredt. Det har også skjedd store framskritt når det gjelder maskinoversettelse.

Som en hovedregel finnes det to typer maskinoversettelsesverktøy; regelbasert og statistikkbasert maskinoversettelse. MT@EC er et eksempel på det siste.

I regelbasert maskinoversettelse styres oversettelsen av en rekke kompliserte lingvistiske grammatikker og regler. Fagmiljøene strides om hvilke metode som gir best resultat. I praksis har man i stadig større grad

³ http://ec.europa.eu/isa/ready-to-use-solutions/mt-ec_en.htm

⁴ <https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation+Benefits>

gått over til statistikkbaserte oversettelsesverktøy som er metoden som benyttes av eksempelvis Google.

Statistikkbaserte maskinoversettelsesverktøy innhenter informasjon fra et minne som består av et corpus. Corpora (flertallsform av corpus) er et sett med tekster som er på et strukturert elektronisk format. Corpora består av parallelle segmenter. Et segment kan eksempelvis være en setning, et kulepunkt eller en overskrift på to språk. Det tekniske formatet på corpora er vanligvis et såkalt TMX-format.

Som navnet tilsier genererer statistikkbaserte maskinoversettelsesverktøy forslag til oversettelse av segmenter basert på statistikk. Verktøyet oversetter segmentet basert på hva som er den mest sannsynlige oversettelsen av segmentet. Dersom det eksempelvis finnes et identisk segment i minnet, vil programmet anslå at sannsynligheten er svært høy for at dette er den korrekte oversettelsen. Programmet gir også forslag til oversettelse dersom det ikke finnes en perfekt match.

2.4 Beskrivelse av MT@EC

MT@EC fungerer ved at såkalte motorer (programmer) oversetter mellom ulike språkkombinasjoner. For ofte brukte språkkombinasjoner bygger EU-kommisjonen direktemotorer. MT@EC har motorer som oversetter direkte mellom engelsk og alle EU-språk og enkelte andre mye brukte språkkombinasjoner. For mindre brukte språkkombinasjoner oversetter MT@EC via engelsk. Dette øker tiden det tar å oversette en tekst, og kan gi en noe dårligere kvalitet på maskinoversettelsen.

Ettersom MT@EC er statistikkbasert, er det avgjørende for kvaliteten på oversettelsen at datagrunnlaget er stort nok. Kvaliteten på MT@ECs maskinoversettelse er bedre, jo større corpus man har målt i antall segmenter. Totalt opplyser oversetterseksjonen i EU om at det i dag er 900 millioner segmenter i MT@ECs corpora. Selv om MT@EC i dag oversetter mellom 24 språk er det svært varierende størrelse på minnet som finnes for ulike språkkombinasjoner. Dette påvirker dermed kvaliteten på oversettelsen mellom ulike språk. Corpora varierer fra 58 millioner segmenter på fransk/engelsk til 2,8 millioner på kroatisk/engelsk.

Dagens norsk bokmål-engelsk corpus i MT@EC består av de oversatte tekstene av EØS-regelverket. Dagens størrelse på det norske corpuset er omtrent 750 000 segmenter. Til sammenligning har det svensk-engelske corpuset en størrelse på om lag 40 millioner segmenter.

På grunn av den begrensede størrelsen på corpuset er foreløpig kvaliteten på oversettelsen til/fra bokmål noe svak. Det er bakgrunnen for at vi i dette oppdraget vurderer om man bør bygge ut tekstgrunnlaget som verktøyet baserer seg på.

Hver gang oversetteravdelingen ferdigstiller en oversettelse av en tekst legges den inn i minnet slik at corpora stadig vokser og kvaliteten på oversettelsene øker. Oversetterseksjonen i EU oversetter imidlertid ikke til norsk. For å bygge opp corpora på norsk må man derfor avlevere corpus til oversetterseksjonen i EU-kommisjonen.

Hva slags type terminologi som benyttes i corpuset, vil også påvirke resultatet av maskinoversettelsen. Dagens corpora i MT@EC er hovedsakelig basert på oversettelser av offisielle EU-dokumenter og EØS-regelverket. Oversettelsene preges dermed av at tekstene i corpuset i stor grad benytter økonomisk og administrativ terminologi.

MT@ECs corpora dekker i dag mange fagområder og domener, men det er foreløpig ingen mulighet til å fokusere på enkeltdomener innenfor visse fagområder. I videreutviklingen av MT@EC er imidlertid et av satsingsområdene å bygge motorer eller programmer som gjør at man i framtiden kan få oversettelser som er rettet mot spesielle fagområder.

2.5 Dagens anvendelser av MT@EC

Det finnes i dag primært tre måter MT@EC anvendes på:

- En webtjeneste for oversettelse av tekst
- Anvendelse i EUs online tjenester/plattformer
- Anvendelse av MT@EC for profesjonelle oversettere

Vi redegjør i det følgende om anvendelsesområdene.

2.5.1 Webtjeneste for oversettelse av tekster

For det første finnes det en webtjeneste hvor man kan laste opp dokumenter på norsk eller et annet EU-språk og få dokumentet oversatt til et valgfritt EU-språk. Det oversatte dokumentet kan lastes ned via webtjenesten eller mottas på epost. Webtjenesten retter seg mot ansatte i EU-institusjoner og etater, samt ansatte og saksbehandlere i offentlige etater i medlemslandene.

Dette verktøyet har som formål å være et hjelpemiddel i det daglige virke til saksbehandlere og andre ansatte i forvaltningen. Verktøyet kan benyttes til å forstå rapporter, strategier og lignende på andre EU-språk. Det kan også være et hjelpemiddel for skriftlig oversettelse av tekster på andre EU-språk, men det vil da kreve etterredigering.

Webtjenesten har også funksjonalitet som gjør at den kan benyttes til å oversette flere dokumenter til flere språk. Ved oversettelse av dokumenter får man opp en kvalitetsindikator på den aktuelle språkkombinasjonen i form av et BLUE-score (man får ikke opp en kvalitetsindikator for den spesifikke teksten).

Verktøyet håndterer en rekke filformat, som Word, Excel, PowerPoint og liknende. Formatering og dokumenttype bevares. Eksempelvis kan man laste opp en PowerPoint-presentasjon og få tilbake en oversatt versjon av presentasjonen i PowerPoint-format, der formatering (skrifttype, farger og liknende) er opprettholdt. Det er i webtjenesten også muligheter for å laste ned oversatte tekster på formater som anvendes av profesjonelle oversettere (TMX-format eller XLIFF-format).

Webtjenesten har også et høyt sikkerhetsnivå for å sikre at konfidensialitet og sikkerhet ivaretas ved oversettelse av dokumenter. Webtjenesten har derfor tilgangskontroll via «EC Authentication Service» (ECAS). Dokumentoverføring skjer enten via sTESTA som er et sikkert privat nettverk mellom offentlig administrasjon i EU separat fra internett eller via internett gjennom en sikker https-kobling.

2.5.2 MT@EC i online tjenester som er finansiert eller støttet av EU

MT@EC er også integrert i ulike online tjenester som retter seg mot offentlig administrasjon, innbyggere og bedrifter i EU. Ved å integrere maskinoversettelse i disse tjenestene kan brukeren benytte tjenestene på sitt eget språk. Tabell 2-1 gir en oversikt over webportaler og nettsteder hvor MT@EC er integrert i dag.

Tabell 2-1: Webportaler og nettsteder hvor MT@EC er integrert

IMI - Internal Market Information System	Portal for ansatte i offentlige virksomheter. Gjennom portalen skal eksempelvis en saksbehandler i Norge kunne kommunisere direkte med en saksbehandler i et annet land på eget språk. MT@EC er introdusert i systemet gjennom oversettelse av tekst i dialogfelt.
SOLVIT	Portal for privatpersoner, som skal bistå med å problemer knyttet til brudd på EØS-rettigheter. MT@EC er introdusert i systemet gjennom oversettelse av tekst i dialogfelt.
Eur-Lex/nLex	Portal som samler alle medlemslandenes nasjonale lovverk. MT@EC er introdusert i systemet gjennom oversettelse av søkeresultater.
CircaBC	Kommunikasjon og informasjon ressurscenter for administrasjon, bedrifter og innbyggere. MT@EC benyttes til oversettelse av dokumenter.
TED	Portal for kunngjøring av offentlige innkjøp i EU. MT@EC er introdusert i systemet gjennom oversettelse av kunngjøringer.

Kilde: DG Translate

MT@EC kan eksempelvis benyttes til å oversette dokumenter som lastes opp eller ned på slike sider, til å oversette tekst i dialogfelt for kommunikasjon mellom brukere eller til å returnere søkeresultater på et spesifisert språk. Foreløpig kan ikke MT@EC anvendes til oversettelse av fulle nettsteder.

2.5.3 Anvendelse av MT@EC i oversetteravdelingen i EU-kommisjonen

I EU-kommisjonen er MT@EC integrert i oversetternes profesjonelle oversetterverktøy slik at forslag fra MT@EC kommer opp parallelt med forslag fra lokalt minne. Det er usikkert om EU-kommisjonen kan tilby en tilsvarende integrasjon for profesjonelle oversettere i de ulike medlemslandene eller om de må basere seg på nedlastning av filer fra webtjenesten som manuelt må importeres i evt. oversettelsesverktøy oversetterne benytter.

2.6 Fremtidige anvendelser av MT@EC i CEF.AT

I videreføringen av MT@EC i CEF.AT vil fokus være på å bygge en sikker og fleksibel maskinoversettelsesplattform av høy kvalitet som kan integreres i EUs webtjenester. Hovedformålet med CEF.AT er å gjøre denne type digitale tjenester flerspråklige. Eksempelvis ønsker man at brukere av ulike digitale tjenester skal få tilgang på informasjon og kommunisere med motparten på morsmålet.

For å videreutvikle MT@EC vil man blant annet ha fokus på:

- Å samle inn tekst i enda større omfang fra ulike kilder for å:
 - øke kvaliteten på oversettelsene
 - bygge oversettelsesmotorer som er skreddersydd ulike fagområder/domener
- Videreutvikle den tekniske løsningen for å forbedre kvaliteten og effektiviteten på oversettelsene.

I dag finnes det bare et generelt corpus i MT@EC som primært inneholder oversettelser av ulike EU-dokumenter. I CEF.AT vil EU-kommisjonen tilby en løsning som gjør at man kan skreddersy oversettelsen til terminologi og teksttyper som er typisk innenfor ulike bruksområder og fagområder. Eksempelvis slik at oversettelsen tilpasses til oversettelse av anbud, juridiske tekster og medisinsk terminologi.

I Tabell 2-2 presenterer vi webtjenester hvor det foreligger konkrete planer om å integrere maskinoversettelse. Når det gjelder ODR lanseres plattformen for norske forbrukere forhåpentligvis tidlig i 2017. BLD har arbeidet med å lansere ODR for norske forbrukere med oversettelse ved hjelp av MT@EC til/fra norsk bokmål. Vi gir mer informasjon om dette arbeidet i neste kapittel.

Tabell 2-2: Webportaler og nettsteder hvor MT@EC vil integreres på sikt som en del av CEF-plattformen

Europeana	Digitalt europeisk bibliotek med et felles flerspråklig aksesspunkt til digitale ressurser
ODP - Open Data Portal	Felles europeisk åpen portal for tilgang til åpne datainfrastrukturer
EESSI - Electronic Exchange of Social Security Information	Plattform for samarbeid mellom 32 land
ODR	Online Dispute Resolution - Online plattform for forbrukerklager
e-justice	En portal som gir et felles aksess til lovverk og muliggjør juridisk samarbeid
SaferInternet	Tjenester for å gjøre internett til et trygt miljø for barn

Kilde: DG Translate

3. Vurdering av behov for MT@EC i statlige virksomheter

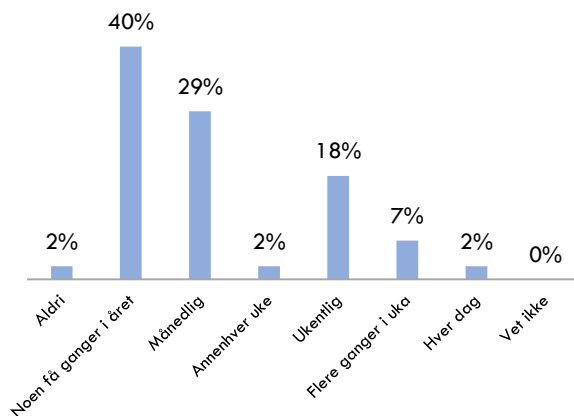
For å vurdere behovet for MT@EC i statlige virksomheter redegjør vi først for statlige virksomheters generelle oversettelsesbehov og ressursbruk knyttet til oversettelse. Vi har også kartlagt hva slags digitale hjelpemidler som benyttes i dag. I lys av dette, samt innspill fra ulike statlige virksomheter, vurderer vi behovet for MT@EC i statlige virksomheter.

3.1 Statlige virksomheters oversettelsesbehov

Departementer, etater og øvrige virksomheter opererer innenfor et bredt spekter av virksomhetsområder og fagområder. Behovet for oversettelse vil dermed variere. I det følgende presenterer vi funn fra vår kartlegging av statlige virksomheters oversettelsesbehov.

Vi har gjennomført en spørreundersøkelse blant statlige ansatte som undersøker deres behov for oversettelser, se Figur 3-1. 27 prosent oppgir at de har behov for å få oversatt tekster og dokumenter ukentlig eller oftere.

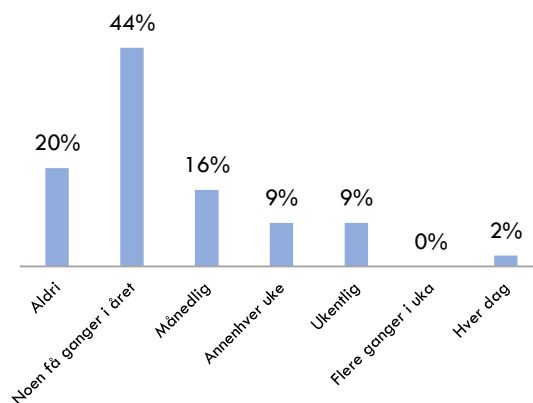
Figur 3-1: Svarfordeling «Hvor ofte har du behov for å få oversatt tekster og dokumenter (f.eks. Word, PowerPoint, Excel) mellom norsk bokmål og et EU-språk i ditt arbeid?», 45 svar



Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

I tillegg til å ha behov for en skriftlig oversettelse, kan ansatte også ha behov for å forstå innholdet i en tekst på et annet språk. Vi har i spørreundersøkelsen spurt respondentene om hvor ofte de har behov for å forstå tekster på andre språk enn engelsk, se Figur 3-2. Ca. 11 prosent av respondentene har behov for å forstå tekster på andre språk enn engelsk ukentlig eller oftere.

Figur 3-2: Svarfordeling «Hvor ofte har du behov for å forstå innholdet i tekster (dokumenter, brev, artikler, eposter, etc.) på andre EU-språk enn engelsk?», 45 svar



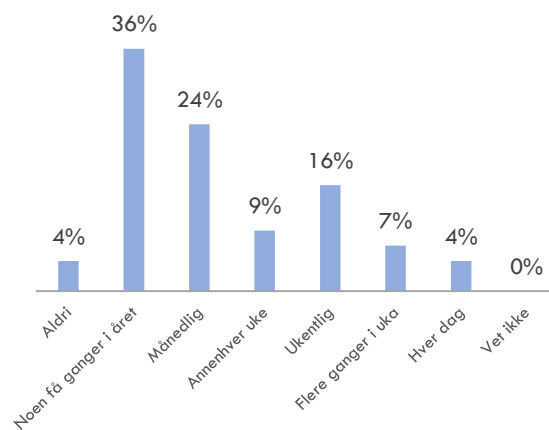
Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

3.1.1 Hvem oversetter tekster?

En stor del av oversettelsene i statlig sektor blir gjennomført i form av innkjøp av oversettelsestjenester fra private leverandører. Omfanget på dette reflekteres av de ulike virksomhetenes ressursbruk til oversettelse som vi redegjør for senere i kapitlet.

Den enkelte ansatte oversetter også i en viss grad tekster selv i forbindelse med saksbehandling eller annet. I spørreundersøkelsen blant ansatte ble respondentene spurt om hvor ofte de oversetter tekster og dokumenter mellom norsk bokmål og et EU-språk. Resultatene vises i Figur 3-3. 27 prosent oppgir at de oversetter tekster selv ukentlig eller oftere.

Figur 3-3: Svarfordeling «Hvor ofte oversetter du selv tekster og dokumenter (f.eks. Word, PowerPoint, Excel) mellom norsk bokmål og et EU-språk i ditt arbeid?», 45 svar



Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

Enkelte virksomheter har interne profesjonelle oversettere som leverer oversettelsestjenester internt i virksomheten og i noen tilfeller også til andre statlige virksomheter. Utenriksdepartementets EØS-avdeling er

eksempelvis en stor leverandør av oversettelsestjenester til departementene. I Tabell 3-1 gir vi en oversikt over de interne oversettere.

Tabell 3-1: Oversettere i statlige virksomheter og type oversettelser som gjennomføres

Virksomhet	Antall årsverk – interne oversettere	Type tekster som oversettes	Type terminologi i tekster	Språk det oversettes til/fra
UD ⁵ - EØS-oversettelse	19	Oversetter rettsakter i form av de delene av EUs regelverk som innlemmes i EØS-avtalen.	Juridisk, teknisk og fagspesifikk terminologi fra en rekke fagområder.	Oversetter fra engelsk til norsk bokmål.
UD - Seksjon for oversettelse	6,5	Oversetter tekster innenfor UDs mange fagfelt. Eksempelvis folkerettslige avtaler (multilaterale og bilaterale avtaler), politiske dokumenter/tekster og UDs strategier og handlingsplaner.	Ulike typer fagterminologi benyttes som reflekterer UDs mange saksfelt (eksempelvis traktat-/avtaleterminologi og terminologi innen juss generelt, kvinners stilling/likestilling, menneskerettigheter, utdanning, migrasjon, klima/miljø, FN og EU)	Oversetter hovedsakelig mellom engelsk og norsk bokmål.
Statens vegvesen	2,5	Oversetter et bredt spekter av tekster, men lengre tekster settes som regel ut til eksterne oversettere. Oversetter eksempelvis regler og forskrifter, veg- og informasjonsskilt, powerpoint-presentasjoner, brev og elektroniske løsninger/søknadsskjema.	Juridisk terminologi, teknisk terminologi, økonomisk og administrativ terminologi, samt fagspesifikk terminologi eks. innen kjøretøyområdet, trafikk miljø og teknologi.	Oversetter fra norsk bokmål til engelsk.
Sjøfartsdirektoratet	2,1	Oversetter med forskriftsendringer fra norsk til engelsk og internasjonalt regelverk fra engelsk til norsk.	Juridisk og teknisk terminologi	Oversetter mellom engelsk og norsk bokmål.
Finanstilsynet	1	Oversetter årsmeldingen, rapporten «Finansielt utsyn», lover, forskrifter og rundskriv til engelsk. Oversettelser av taler og presentasjoner kan også forekomme.	Økonomisk og juridisk terminologi	Oversettelser til engelsk
Stortinget	0,5	Politiske taler og tekster til Stortingets hjemmesider	Generell terminologi	Oversettelse fra norsk bokmål til engelsk
Totalt 31,6 årsverk				

⁵ Utenriksdepartementet

3.1.2 Hvilke tekster oversettes?

En stor del av oversettelsene som gjennomføres av de interne oversetterne er av juridiske tekster. Herunder er en stor del av oversettelsene til norsk oversettelser av EU-regelverk. Utenriksdepartementets enhet for EØS-oversettelse oversetter til norsk bokmål de deler av EUs regelverk som innlemmes i EØS-avtalen. Dette er på vegne av departementene som forvalter regelverket. Regelverket består av rettsakter i form av forordninger, direktiver og beslutninger.

Oversettelser av juridiske tekster har høy vanskelighetsgrad og krever at man benytter juridisk terminologi og traktatterminologi. I tillegg vil terminologien som kreves avhenge av fagområdet regelverket gjelder innenfor slik at det også vil være store innslag av teknisk og annen fagspesifikk terminologi fra en rekke fagområder.

UDs seksjon for oversettelse oversetter folkerettslige avtaler til norsk bokmål. Dette er internasjonale bilaterale eller multilaterale avtaler mellom stater og eller mellom stater og internasjonale organisasjoner. Oversettelser av folkerettslige avtaler er krevende og krever høy grad av presisjon i språket.

Sjøfartsdirektoratet oversetter forskriftsendringer som gjelder sjøfartsområdet fra norsk bokmål til engelsk og oversetter internasjonalt regelverk fra engelsk til norsk bokmål. Statens vegvesen oversetter også juridiske tekster i form av regler og forskrifter på transportområdet.

I tillegg til juridiske tekster oversetter de interne oversetterne også tekster som hjemmesider, politiske dokumenter/tekster, strategier og handlingsplaner, skjemaer og veiledere og annet. Det er også en rekke tekster av mer teknisk art som oversettes. Statens vegvesen oversetter eksempelvis tekniske håndbøker. Forskriftene på Sjøfartsområdet vil også ha betydelig innslag av teknisk terminologi.

Private leverandører oversetter alt fra offentlige rapporter på ulike fagområder, forskrifter, stortingsmeldinger, vedtak, kontrakter og avtaler til dommer, pressemeldinger, førerkortbekreftelser, maler, policy-dokumenter, presentasjoner og andre HR-/HMS-dokumenter. For utdanningssektoren oversettes det også eksamener, studieplaner og liknende.

3.1.3 Hvilke EU-språk oversetter statlige virksomheter til/fra?

Intervjuene med de interne oversetterne viser at oversettelsene som gjennomføres hovedsakelig er mellom norsk bokmål og engelsk. Tidligere oversatte de interne oversetterne i staten også til andre språk,

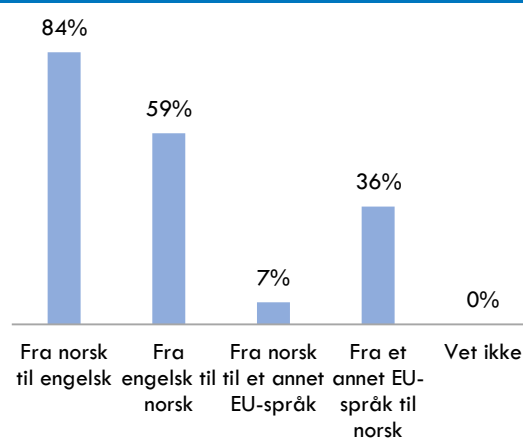
men dette har man i all hovedsak gått bort fra som følge av ressursbegrensninger.

Hovedtyngden av eksterne innkjøp av oversettelser er mellom norsk bokmål og engelsk. Utover dette får de private oversetterne også noen oppdrag som gjelder oversettelser til/fra språk som italiensk, tysk, spansk, gresk, polsk, fransk, nederlandsk, portugisisk, dansk, svensk, ungarsk, kroatisk m.fl. Ingen av disse språkene utgjør en betydelig mengde.

I spørreundersøkelsen blant ansatte opplyser hovedparten at respondentene at de har behov for å oversette tekst mellom norsk bokmål og engelsk i sitt arbeid, se Figur 3-4. Videre oppgir 36 prosent av respondentene at de har behov for å oversette fra et annet EU-språk (enn engelsk) til norsk bokmål. Disse respondentene oppgir at de har behov for å oversette tekst til norsk bokmål fra fransk, tysk, nederlandsk, spansk, italiensk, dansk og svensk.

Kun 7 prosent har behov for å oversette fra norsk bokmål til et annet EU-språk (enn engelsk) i sitt arbeid.

Figur 3-4: Svarfordeling «Mellom hvilke EU-språk har du behov for å få oversatt tekst i ditt arbeid?», 44 svar



Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

I spørreundersøkelsen blant virksomhetene svarer de fleste at de primært har behov for å oversette mellom engelsk og norsk bokmål. Andre språk som virksomhetene har behov for å oversette til/fra er nordiske språk, polsk, baltisk, tysk, fransk, spansk og italiensk og generelt til språk som benyttes i nasjoner Norge samhandler mye med.

Noen eksempler på innspill fra virksomheter:

- NFD melder blant annet om behov for oversettelse til språkene til ikke-engelskspråklige kystland, dvs. særlig til fransk, spansk, portugisisk, italiensk, samt språkene på Balkan og i Baltikum.

- Kripos har behov for oversettelse til mindre språk i Øst-Europa.
- NAV har behov for oversettelse av informasjon til språk som er utbredt blant deres brukere.
- NOKUT har behov for oversettelse til engelsk, tysk, fransk og noe spansk.
- UDI melder at maskinoversettelse trolig primært er relevant for oversettelse til andre EU-språk enn engelsk.
- KLD har behov for oversettelse til spansk, portugisisk, fransk og engelsk.
- Havarikommisjonen har primært behov for oversettelse til engelsk, men også språk som tysk og polsk for å imøtekomme involverte parter fra aktuelle land.
- NKOM antar at de mest aktuelle språkene er tysk, fransk, spansk og italiensk.

3.2 Ressursbruk knyttet til oversettelse i statlige virksomheter

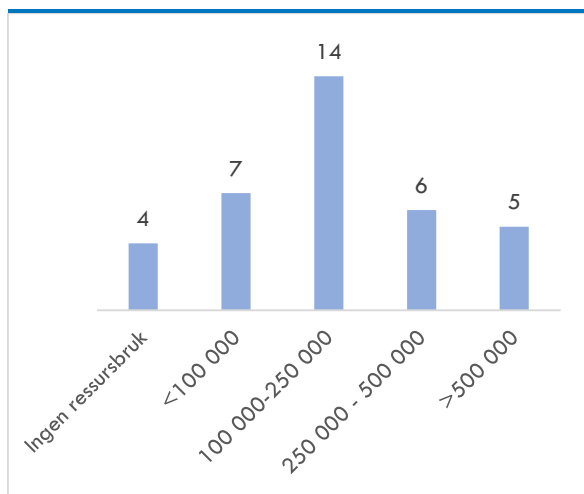
I det følgende gir vi estimater på ressursbruk til oversettelser i statlig sektor basert på våre undersøkelser.

3.2.1 Kostnader knyttet til innkjøp av oversettelsestjenester

En stor del av oversettelsene i statlig sektor utføres av private aktører. Vi har i intervjuer og i spørreskjemaet sendt ut til statlige virksomheter kartlagt ressursbruk til eksterne leverandører av oversettelse.

Fordelingen i ressursbruken hos de statlige virksomhetene hvor vi har mottatt kostnadsopplysninger fremgår av Figur 3-5.

Figur 3-5: Fordeling i utvalget per intervall med ressursbruk på private oversettere i kr eks mva., 35 statlige virksomheter



⁶ <https://www.regjeringen.no/no/tema/arbeidsliv/Statlig-arbeidsgiverpolitikk/statens-sentrale-tjenestemannsregister/id477017/>

Kilde: Spørreundersøkelse blant statlige virksomheter

Totalt har utvalget av statlige virksomheter utgifter på 13 millioner kroner til eksterne oversettere. I gjennomsnitt er utgiften per virksomhet omtrent 370 000 kr i året.

Figuren viser at 85 prosent av utvalget har en ressursbruk under 250 000 kr årlig til eksterne oversettere. Enkelte virksomheter skiller seg ut ved at de har store utgifter til eksterne oversettere. Blant virksomhetene som har utgifter over 500 000 kr finner vi UD, JD, SSB, FIN og BLD. UD's utgifter består hovedsakelig av bistand til oversettelse av EU-regelverk, som utgjør ca. 4,5 millioner kroner årlig. Disse oversettelsene gjennomføres på vegne av de ulike departementene, og representerer dermed ikke UD sitt forbruk av oversettertjenester som sådan. JD har også store utgifter til oversettelse (ca. 1,2 mill. kroner årlig). Oversettelsene som gjennomføres i JD består av ulike offentlige dokumenter utarbeidet av JD eller dokumenter som kommer fra EU (ikke EU-regelverk), samt rettsanmodninger og brev fra andre land.

3.2.2 Ressursbruk til interne oversettelser

I tillegg til kostnader knyttet til de eksterne oversetterne har de statlige virksomhetene kostnader i form av ressursbruk til interne oversettere i staten. I Tabell 3-1 gir vi en oversikt over antall årsverk for intern (profesjonell) oversettelse.

Våre undersøkelser anslår at det årlig medgår ca. 32 årsverk i staten for oversettelse som gjennomføres av de interne oversetterne. UD har klart flest årsverk med 25,5 årsverk totalt. I tillegg har Sjøfartsdirektoratet og Statens vegvesen rundt 2 årsverk hver tilknyttet intern oversettelse.

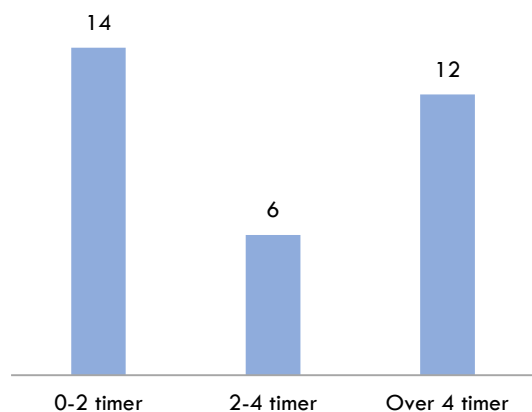
Ansatte i staten oversetter også til en viss grad tekster i sitt daglige arbeide. I Figur 3-6 under viser vi resultatene på spørreundersøkelsen blant ansatte hvor respondentene ble spurt om å anslå sin tidsbruk knyttet til å oversette tekster.

Oppgitt gjennomsnittlig tidsbruk per måned på oversettelse er i utvalget 7 timer. I utvalget varierer oppgitt ressursbruk fra 0 til 65 timer per måned. Totalt i utvalget oppgis en ressursbruk knyttet til oversettelser på ca. 2600 timer i året.

Det er grunn til å tro at utvalget representerer et øvre estimat på ressursbruk knyttet til oversettelser. Det er ifølge Statens sentrale tjenestemannsregister 153 357 ansatte i de statlige virksomhetene per 1.10.2014.⁶ Den totale ressursbruken blant statlige ansatte knyttet

til oversettelser kan dermed være betydelig, men det hersker stor usikkerhet knyttet til størrelsen på ressursbruken.

Figur 3-6: Fordeling utvalg: «Hvor mange timer vil du anslå at du bruker på å oversette tekster i løpet av en måned?»», 32 svar



Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

3.2.3 Samlet anslag på ressursbruk knyttet til oversettelse

Basert på resultatene fra spørreundersøkelsen blant de ansatte og informasjonen vi har mottatt fra virksomhetene, anslår vi samlet ressursbruk i statlig sektor knyttet til oversettelse. Estimatenes presenteres i Figur 3-7.

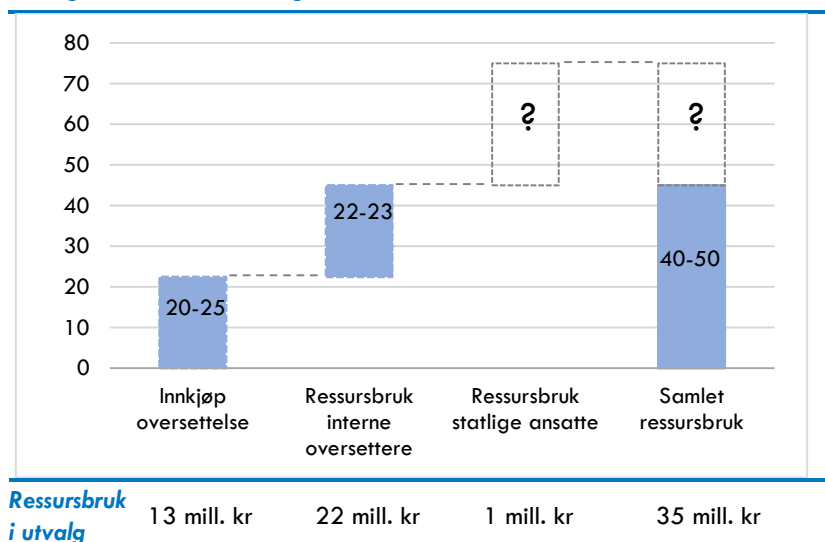
Vi estimerer total ressursbruk knyttet til innkjøp av oversettelsestjenester basert på Amestos årlige omsetning knyttet til oversettelser for statlige virksomheter og anslått markedsandel.

For å verdsette ressursbruk til interne oversettere legger vi til grunn en kostnad per årsverk på 684 190 kr.⁷ Verdien på ressursbruken knyttet til de interne oversetterne vi har kartlagt er da 22 mill. kr. Vi tar imidlertid høyde for at det kan være interne oversettere i staten som vi ikke har kartlagt i våre undersøkelser og vi anslår derfor at ressursbruken knyttet til interne oversettere er mellom 22 og 23 millioner kroner.

Når det gjelder tidsbruken knyttet til oversettelser som gjennomføres av ansatte i statlige virksomheter som ikke er profesjonelle er det betydelig usikkerhet. Vi gir dermed ikke et konkret anslag på denne størrelsen. Det er likevel grunn til å tro at ressursbruken er betydelig. I utvalget vårt oppgir respondentene at de totalt bruker 2600 timer i året. Vi verdsetter denne tiden med en timepris på 400 kr.⁸ Total ressursbruk til oversettelse i utvalget er dermed estimert til ca. 1 million kr. Dette tilsvarer en kostnad på ca. 30 000 kr per ansatt i utvalget årlig.

Basert på disse estimatene er samlet ressursbruk i statlig sektor estimert til å være 40-50 millioner kroner dersom vi ser bort fra ressursbruk for oversettelser som utføres av vanlige ansatte.

Figur 3-7: Estimert årlig ressursbruk til oversettelse i mill. kr eks mva



⁷ Kostnad per årsverk er estimert basert på gjennomsnittlig inntekt blant statlige ansatte på 526 300 (SSB tabell 08702) med et 30 prosent påslag for sosiale utgifter og overhead.

⁸ Vi legger til grunn en årsverkskostnad 684 190 kr og at et årsverk tilsvarer 1695 timer (I Norge har mange

arbeidstakere har tariffestet 37,5 timers arbeidsuke og 5 ukers ferie. I et år med 365 dager, 25 feriedager, 104 helgedager, 10 andre røde dager og 7,5 timers arbeidsdager tilsvarer dermed et årsverk 1695 timer).

3.3 Bruk av digitale hjelpemidler til oversettelse

Dagens bruk av digitale hjelpemidler for oversettelse har også betydning for behovet for MT@EC. Dersom det allerede finnes gode oversettelsesverktøy kan det bety at behovet for MT@EC er mindre. Vi presenterer derfor i det følgende hva slags digitale verktøy for oversettelse som benyttes av interne oversettere og øvrige ansatte i statlige virksomheter.

3.3.1 Interne oversettere

I intervjuene med de interne oversetterne har vi kartlagt hva slags digitale hjelpemidler for oversettelse som benyttes i dag. De fleste interne profesjonelle oversetterne i staten benytter såkalte CAT-verktøy («Computer-Assisted-Translation»-verktøy). Dette er et dataprogram som støtter og faciliterer oversetterens arbeid. I Tabell 3-2 presenterer vi en oversikt over hvilke CAT-verktøy som benyttes av de interne oversetterne.

CAT-verktøyene har en rekke funksjoner som gjør arbeidet med oversettelse mer effektivt.

Et CAT-verktøys viktigste funksjoner inkluderer:

- Segmentering av teksten i setninger og presentasjon av segmentene på en egnet form.
- Lagring av segmentene i et minne.
- Søkbart minne og/eller automatiske forslag til oversettelse av segmenter (typisk setninger) basert på minne

CAT-verktøyet inkluderer også en funksjonalitet i form av maskinoversettelse ettersom det presenteres forslag til oversettelse basert på minne. Dersom segmentet allerede finnes i minnet presenteres en 100 prosent treff og evt. får man opp andre forslag til oversettelse basert på minnet som er basert på delvis treff. Det skiller seg imidlertid fra maskinoversettelse

ved at oversetteren aktivt tar stilling til hvert enkelt segment.

Tabell 3-2: Oversikt over digitale hjelpemidler som benyttes av de interne oversetterne

Virksomhet	Digitale hjelpemidler til oversettelse (CAT-verktøy)
UD - EØS-oversettelse	SDL Trados Studio
UD - Seksjon for oversettelse	MultiTrans Prism
Statens vegvesen	Memsources
Sjøfartsdir.	SDL Trados Studio
Finanstilsynet	Benytter ingen verktøy
Stortinget	Benytter ingen verktøy

Kilde: Intervjuer med de interne oversetterne

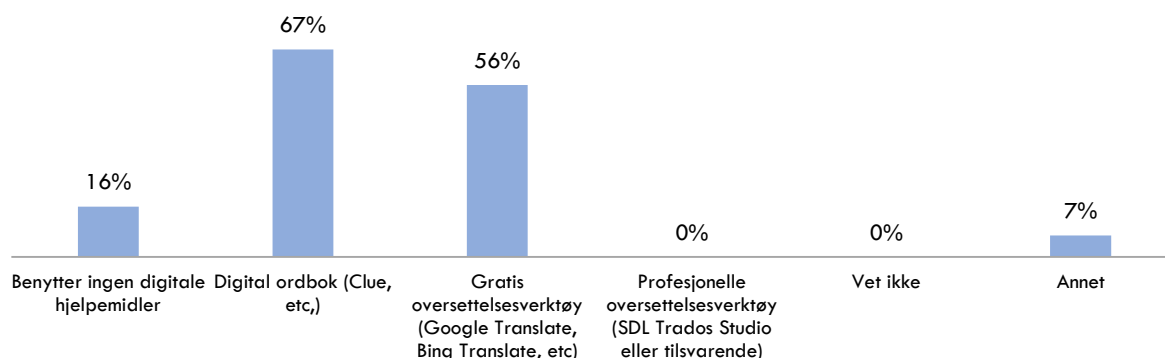
De aller fleste eksterne (private) leverandører benytter seg også av CAT-verktøy. I de tilfellene hvor en statlig enhet benytter seg av interne samt eksterne oversettere (f.eks. Utenriksdepartementet og Statens vegvesen), kan minnene enkelt deles mellom interne og eksterne oversettere. Det fører til at begge parter kan dra nytte av det hverandre gjør.

Minnene lagres i et format som enkelt kan konverteres mellom de ulike CAT-verktøyene. Man risikerer derfor ikke å miste noe av det man har jobbet med tidligere, om man skulle bytte verktøy eller samarbeide med eksterne leverandører som bruker et annet verktøy.

3.3.2 Ansatte i statlige virksomheter

Vi har også kartlagt i hvilken grad ansatte i statlige virksomheter benytter digitale hjelpemidler. I spørreundersøkelsen ble respondentene blant annet spurt om hvilke digitale hjelpemidler de benyttet for å oversette tekster, se Figur 3-8 for svarfordeling.

Figur 3-8: Fordeling utvalg: «Hvilke digitale hjelpemidler benytter du i dag når du oversetter tekster?», 43 svar



Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

Det er ingen av respondentene som oppgir at de benytter profesjonelle oversettelsesverktøy (CAT-verktøy). Bruken av Google Translate og tilsvarende verktøy er relativt utbredt og over halvparten oppgir at de benytter slike verktøy. Google Translate kan potensielt være problematisk dersom tekster med sensitive opplysninger oversettes i verktøyet. Vi har imidlertid ikke kartlagt hva slags type tekster respondentene oversetter med Google Translate. I spørreundersøkelsen blir respondentene også spurt om hvor enig de er i påstanden «Jeg benytter i dag ofte Google Translate eller tilsvarende online oversettelsestjenester i mitt arbeid». Her oppgir om lag halvparten at de er «Litt enig» eller «Helt enig» i denne påstanden.

Bruken av digital ordbok er også utbredt og 67 prosent av respondentene oppgir at de benytter digital ordbok.

3.4 Behov for oversettelse til/fra norsk i EUs webportaler

Som tidligere nevnt inngår MT@EC i mange europeiske samhandlingsverktøy som brukes i dag, hovedsakelig av offentlige myndigheter. I videreføringen av MT@EC er strategien at MT@EC skal integreres i stadig flere av EUs digitale tjenester.

Foreløpig oversettes det ikke til norsk ved hjelp av MT@EC i disse plattformene. Vi har i intervjuer og spørreundersøkelser kartlagt hvor MT@EC benyttes i dag og i hvilken grad det er behov for oversettelse til/fra norsk bokmål i ulike EU-portaler.

ODR er eksempel på en digital tjeneste hvor MT@EC benyttes til å oversette tekst mellom ulike EU-språk. Man arbeider nå med å gjøre ODR tilgjengelig for norske forbrukere. I den forbindelse vil man få norsk inn som språk i MT@EC. Siden brukerne av plattformen primært er privatpersoner, ble det sett på som viktig at de kan bruke norsk, og ikke er henvist til for eksempel engelsk eller dansk. BLD forvalter ODR-regelverket i Norge. For mer informasjon om arbeidet med å innarbeide maskinoversettelse til norsk i ODR, se Tabell 3-3.

Hvor godt ODR-løsningen fungerer for norske forbrukere vil til dels avhenge av kvaliteten på oversettelsen til norsk. Foreløpig har en utfordring vært at det norske corpuset ikke dekker terminologi i dagligspråk som f.eks. ordene «løgn» eller «lure». ODR-løsningen har dermed et stort behov for at MT@EC-corpuset utvides og at dagligspråk og forbrukerområdet dekkes i enda større grad i corpuset.

Tabell 3-3: Caseeksempel: Bruk av MT@EC i ODR

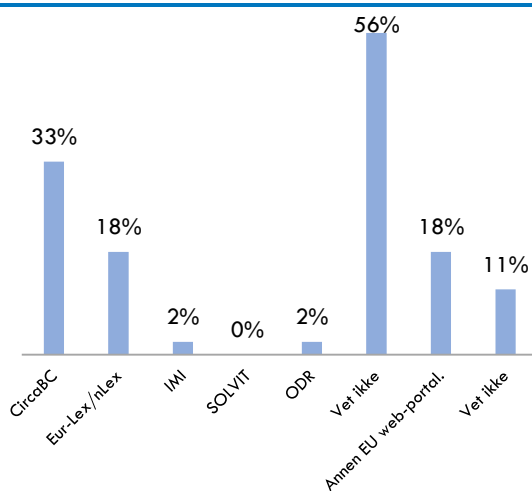
Hva er ODR?	<p>ODR er en nettbasert plattform for forbrukerklager utviklet i EU. ODR gjør at forbrukere som har et problem knyttet til kjøp av en vare eller tjeneste online fra en næringsdrivende i et annet land, kan løse konflikten utenfor domstolene. ODR fungerer ved at forbrukeren fyller ut et online klageskjema. Plattformen identifiserer deretter rett klageorgan og sørger for videre kommunikasjon mellom partene og klagenemnda. Plattformen, som åpnet 15. februar 2016, foreligger på alle de offisielle EU-språkene.</p> <p>Lenke til ODR-plattformen: https://webgate.ec.europa.eu/odr/main/index.cfm?event=main.home.chooseLanguage</p>
Når vil ODR bli tilgjengelig for norske forbrukere?	<p>Når EU-regelverket som regulerer plattformen blir tatt inn i EØS-avtalen, vil plattformen også foreligge på norsk. Norske brukere vil da få tilgang, både til plattformen, og til andre lands klageorganer. Fra 9. januar 2016 er det slik at alle EU-land har plikt til å ha et rimelig og godt klagebehandlingstilbud for nesten alle typer varer og tjenester. Når Norge kobler seg på EU-systemet (forhåpentligvis tidlig i 2017) vil norske forbrukere dra nytte av systemet.</p>
Hvem forvalter ODR?	<p>BLD forvalter ODR-regelverket i Norge, mens Forbruker Europa vil bli formelt oppnevnt som kontaktpunkt for ODR når Norge kobler seg på EU-systemet.</p>
Hvem er brukere?	<p>Målgruppen for ODR er forbrukere, men plattformen vil også benyttes av næringsdrivende og klageorganene.</p>
Hvordan er MT@EC integrert i ODR?	<p>MT@EC benyttes til maskinoversettelse av fritekstfeltene og til oversettelse av dokumenter som lastes opp på siden. Dersom eksempelvis en norsk forbruker ønsker å klage på en vare eller tjeneste kjøpt på nett av en polsk næringsdrivende, kan den norske forbrukeren velge norsk som språk på plattformen, skrive inn informasjon og motta informasjon, for eksempel utfyllende forklaringer, på norsk. Vedlegg til saken, inkludert pdf-filer blir også oversatt mellom språkene. Dette gjøres ved hjelp av MT@EC. Dersom forbrukeren ønsker resultatet av nemndas behandling oversatt til norsk gjøres det av oversettere.</p>
Hvorfor oversettelse til/fra norsk i ODR?	<p>MT@EC inngår i mange europeiske samhandlingsverktøy som brukes i dag, hovedsakelig av offentlige myndigheter. Arbeidet med å få norsk inn som språk i MT@EC i ODR ble startet da arbeidet med gjennomføring av EU-forordning om nettbasert tvisteløsning ble påbegynt. Siden brukerne av plattformen primært er privatpersoner, ble det sett på som viktig at de kan bruke norsk, og ikke er henvist til for eksempel engelsk eller dansk.</p>
Hva har man gjort for å oversette til norsk i ODR?	<p>For å etablere en oversettelse til norsk i ODR baserte man seg i utgangspunktet på det eksisterende corpuset på norsk bokmål-engelsk som hovedsakelig inneholder oversettelser av EØS-regelverket. Kvaliteten på maskinoversettelse til/fra norsk er av EU vurdert som god nok til at verktøyet kan brukes i plattformen.</p> <p>Norge (og Island) vil oversette den statiske teksten i plattformen til eget språk. Dette arbeidet gjøres i Norge av Forbruker Europa. Totalt har arbeidet med å manuelt oversette tekster kostet rundt 200 000 kr.</p>
Hvilke utfordringer har man hatt knyttet til oversettelse til norsk i ODR?	<p>Siden corpuset til MT@EC er basert på juridisk tekst, vil ikke oversettelse av dagligspråk/muntlig språk, som gjerne forekommer i forbrukerklager, fungere så bra. Forbrukere og en stor del av næringsdrivende som skal bruke plattformen vil ventelig benytte en god del muntlig språk i fritekstfeltene. Foreløpig inneholder det norske corpuset ikke terminologi som dekker dagligspråk som f.eks. ordene «løgn» eller «lure». Dette gjør at kvaliteten på oversettelsen blir relativt dårlig. Verktøyet fungerer bedre for klage/meglingsorganenes bruk, men det er heller ikke alle de juridiske termene de benytter som finnes i det eksisterende corpuset.</p> <p>Det vurderes fortløpende, både for norsk og for en del EU-språk om det bør settes inn tiltak for å forbedre kvaliteten, som for eksempel innlasting av ordbøker, sektorspesifikke tekster osv., men det kan ta tid. De som har ansvaret for MT@EC i EU-kommisjonen har også lagt til grunn at oversettelser av mindre volumer fagtekst, ikke vil bidra i særlig grad til forbedring av systemet.</p>

Kilde: Intervjuer og skriftlige innspill fra BLD og Forbruker Europa

Vi har i spørreundersøkelsen kartlagt i hvilken grad det er behov for oversettelse til norsk bokmål i EUs webportaler. Se Figur 3-9 for svarfordeling.

Behovet for oversettelse til norsk bokmål vurderes av respondentene å være størst i CircaBC og Eur-Lex/nLex. Dette henger trolig sammen med at disse webportalene er mest brukt av de som er oppgitt. Henholdsvis 40 prosent og 20 prosent av utvalget oppgir at de benytter henholdsvis CircaBC og Eur-Lex/ nLex/.

Figur 3-9: Fordeling utvalg: «I hvilke av EUs webportaler har du behov for å få oversatt informasjon mellom norsk bokmål og EU-språk?», 45 svar



Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

Få nevner at de har behov for oversettelse til norsk bokmål i de øvrige nettstedene/portalene. Andre webportaler hvor respondentene oppgir at det kunne vært nyttig med oversettelse til norsk bokmål er i Food Additive Database, European Food Safety Authority, IT-Tool, River Information Services (RIS), Trade Control and Expert System (TRACES), Food and Feed Safety Alerts (RASFF).

I Tabell 6-1 i vedlegget har vi oppsummert virksomhetenes innspill når det gjelder bruk av EUs webportaler og behovet for oversettelse til/fra norsk bokmål i disse.

Virksomhetene benytter EUs webportaler og nettsteder i ulik grad. Av de mer generelle webportalene i EU nevner flere virksomheter at de benytter Eur-Lex og CircaBC. TED (Tenders Electronic Daily) benyttes også av enkelte av virksomhetene. I tillegg nevnes en rekke databaser, webportaler og nettsteder som er mer spisset mot virksomhetenes fagområder. Forbrukerombudet benytter eksempelvis Consumer Protection Cooperation System (CPCS) til behandling av europeiske forbrukersaker, mens Havarikommissjonen benytter felles europeiske

databasesystemer innenfor de ulike transportområdene.

Når det gjelder behovet for oversettelse til/fra norsk bokmål i disse webportalene/nettstedene er tilbakemeldingene noe blandet.

Brønnøysundregistrene peker på at det er behov for oversettelse til/fra norsk i IMI. De tror det kan bidra til at det blir delt flere tekster og dokumenter mellom statlige virksomheter i EU ved hjelp av IMI. Videre peker de på at det også er behov for maskinoversettelse til/fra norsk i øvrige EU-portaler som de forvalter som eksempelvis Your Europe.

NFD forvalter IMI i Norge. De vurderer at det ikke er et stort behov for oversettelser til norsk i IMI i dag, men at nytten av oversettelser vil være avhengig av kvaliteten på oversettelsene. De forventer at Europakommisjonen vil benytte IMI på stadig flere områder. Videre vurderer de at økt bruk av IMI kan øke behovet for oversettelser til/fra norsk. NFD påpeker også at notifikering av nasjonalt regelverk for varer (TRIS) og tjenester (IMI) innebærer oversettelse av norske lover/forskrifter til engelsk som muligens kunne blitt mindre ressurskrevende med maskinoversettelse.

Lovdata ser flere bruksmuligheter for MT@EC og tror det kan bidra til å gjøre norsk regelverk tilgjengelig på andre språk enn norsk. De påpeker at norsk regelverk i liten grad er tilgjengelig på andre språk og at vi heller ikke er tilstede på nLex sidene. De mener det bør være et mål for Norge å kunne presentere norsk regelverk på ulike språk gjennom innpass på nLex eller på andre måter.

Forbrukerombudet mener at det hadde vært til stor hjelp dersom MT@EC hadde vært implementert i Consumer Protection Cooperation System (CPCS) som er saksbehandlingssystemet for behandling av europeiske forbrukersaker.

UDI benytter også i stor grad ulike EU-portaler/nettsteder (eksempelvis Eur-Lex, CIRCABC, CURIA og nettsiden til EU-kommisjonen). De vurderer at MT@EC vil være nyttig ved bruk av disse portalene.

Datatilsynet benytter ulike portaler for samarbeid mellom personvernmyndigheter og arbeidsgrupper innen disse. Hovedsakelig er det CircaBC som benyttes. Materiale som legges ut der foreligger for det mest på engelsk, men i noen grad også på tysk og fransk. De vurderer at et slikt verktøy med fordel kunne vært implementert i de samarbeidsportalene som vil bli etablert som følge av det nye personvernregelverket, som skal implementeres i Europa i 2018.

Kripos mener at maskinoversettelse i politiets egne saksbehandlingsverktøy for internasjonale saker og etterretningsinformasjon kunne vært en hjelp for brukerne av disse systemene, og ville ha bidratt til mer effektiv behandling av informasjon og forespørsler. Dette forutsetter imidlertid at systemet ivaretar politiets strenge krav til informasjonssikkerhet.

De øvrige virksomhetene er enten usikre på om det er behov for maskinoversettelse til norsk bokmål eller vurderer at det ikke er det. Mange argumenterer med at de får den informasjonen de trenger på engelsk eller et nordisk språk og at det dermed ikke er behov for oversettelse til norsk bokmål.

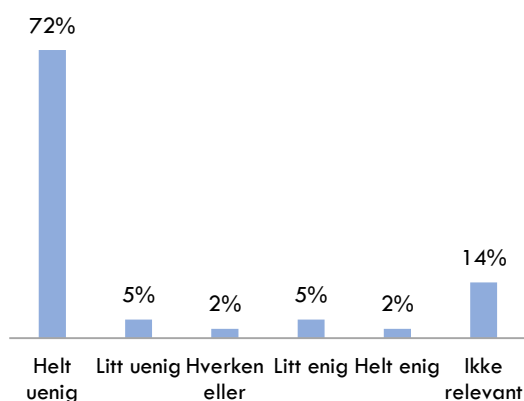
Samlet sett synes det å være behov for oversettelse til/fra norsk bokmål i en del tilfeller. Det er spesielt CircaBC og nLex/Eur-Lex som peker seg ut som webportaler hvor behovet er stort. Det er potensielt også behov for oversettelse til/fra norsk bokmål i enkelte webportaler som er spesifikke til virksomhetene.

3.5 Behov for oversettelse til/fra nynorsk i MT@EC

I intervjuer og spørreundersøkelsen har vi også kartlagt om det er behov for oversettelse til/fra nynorsk og EU-språk. Målloven krever at minst 25 prosent av alle tekster i offentlig forvaltning skal være på nynorsk.

I spørreundersøkelsen har vi kartlagt omfanget på oversettelser mellom nynorsk og EU-språk. I Figur 3-10 viser vi resultatene.

Figur 3-10: Fordeling utvalg: «Hvor enig er du i påstanden? «Jeg oversetter ofte mellom andre EU-språk og nynorsk i mitt arbeid.»», 43 svar



Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

Kun 7 prosent av utvalget er litt enig eller helt enig i påstanden «Jeg oversetter ofte mellom andre EU-

⁹ <http://nynodata.no/index.html>

språk og nynorsk i mitt arbeid»). Dette indikerer at omfanget på oversettelser mellom EU-språk og nynorsk er beskjedent.

Vi har i intervjuer og spørreskjema til virksomhetene også belyst dette spørsmålet. Samtlige virksomheter vurderer at de ikke har behov for oversettelse til nynorsk eller at behovet er svært begrenset. Det er imidlertid enkelte unntak. Helse- og omsorgsdepartementet oppgir at muligheten til å oversette mellom EU-språk og nynorsk er «interessant». Nokut vurderer også at det eksisterer et behov for maskinoversettelse mellom EU-språk og nynorsk. NFD vurderer at det kan være aktuelt å benytte maskinoversettelse mellom EU-språkene og nynorsk for å oppfylle nynorskkravet. De påpeker imidlertid at kravet også oppfylles på andre måter enn ved maskinoversettelse.

De fleste melder om at det viktigste er muligheten til oversettelse mellom EU-språkene og norsk bokmål – så oversettes tekstene videre til nynorsk i den grad det er nødvendig for å oppfylle målloven.

Amesto Translations oppgir at oversettelser fra bokmål til nynorsk er det vanligste etter oversettelser mellom engelsk og bokmål. Omfanget av oversettelser mellom engelsk og nynorsk er så godt som fraværende.

Det finnes allerede maskinoversettelsesverktøy på markedet som oversetter fra bokmål til nynorsk. Nyno er et eksempel på et slikt verktøy.⁹ At denne type verktøy eksisterer kan redusere behovet for et verktøy som MT@EC som oversetter mellom EU-språk og nynorsk. Det vil være mulig å først oversette tekst til bokmål ved hjelp av MT@EC og deretter til nynorsk ved behov. En smidig maskinoversettelse fra EU-språk til nynorsk vil imidlertid i så fall kreve en integrasjon mellom MT@EC og Nyno eller tilsvarende verktøy.

Samlet sett vurderer vi at behovet for oversettelse mellom nynorsk og EU-språk er begrenset.

3.6 Samlet vurdering av behovet for MT@EC

Vi vurderer i det følgende behovet for oversettelse til norsk MT@EC i lys av undersøkelsene vi har gjennomført i virksomhetene.

Vi fokuserer på behovet for å oversette mellom EU-språk og norsk ettersom problemstillingen for oppdraget er behovet for å satse på oversettelse til norsk i MT@EC. MT@EC kan imidlertid også gi gevinster for statlige virksomheter ved at de kan benytte verktøyet til å oversette dokumenter på ulike EU-språk til engelsk.

3.6.1 Profesjonelle oversettere i statlige virksomheter

Interne oversettere benytter hovedsakelig allerede CAT-verktøy. MT@EC kan gi noe gevinst ved at oversetterne får opp flere forslag til oversettelse av ord og setninger. Dersom kvaliteten på forslagene basert på MT@EC er høyere enn kvaliteten på forslagene basert på minne, vil MT@EC gi en gevinst for profesjonelle oversettere. Hvorvidt MT@EC gir oversetterne gevinster vil dermed avhenge av kvaliteten på oversettelsen MT@EC gir og hvor stort minne virksomheten har fra før.

Hvilken terminologi oversetteren har behov for er også av betydning. Generelt er det viktig at MT@ECs corpora inneholder fagterminologien som oversetterne benytter. Dagens corpora inneholder en oversettelse av EØS-regelverket. Fagterminologien er dermed i stor grad juridisk og vil trolig ha begrenset verdi for oversettelser av taler og lignende. Flere av virksomhetene som Utenriksdepartementets seksjon for oversettelser og delvis Utenriksdepartementets EØS-seksjon vurderer at deres tekstminne er rikt nok og at MT@EC ikke vil gi forslag til oversettelser bedre enn de forslagene de allerede får basert på deres egne tekstminner.

Sjøfartsdirektoratet og Statens vegvesen som har mindre oversettermiljø vurderer på den andre siden at det vil kunne gi gevinster for dem å få tilgang på oversettelser gjennomført i andre virksomheter som eksempelvis oversettelsene av EØS-regelverket.

For at MT@EC skal gi oversetterne gevinster er de interne oversetterne klare på at det vil kreve en effektiv integrering i dagens CAT-verktøy. I utgangspunktet har EU-kommisjonen signalisert at oversetterne må laste ned TMX-filer av det enkelte dokumentet i webtjenesten som må importeres i CAT-verktøyet. Oversetterne er skeptiske til en slik løsning, ettersom dette vil gi en tungvint integrering av MT@EC i deres arbeidsstrømmer. Helst ønsker de en løsning tilsvarende den oversetteravdelingene i EU

benytter hvor det er en direktekobling mellom MT@EC og CAT-verktøyene.

Det er også generelt sprikende holdninger til maskinoversettelse blant de interne oversetterne. Flere mener at det er direkte uegnet for den type oversettelser de gjennomfører. UDs seksjon for oversettelser mener verktøyet vil gi få eller ingen gevinster for dem, ettersom erfaringen er at etterredigeringsarbeidet etter bruk av slike verktøy krever vel så mye ressurser som å oversette teksten fra bunnen av. De påpeker at man for å oversette en tekst må bli kjent med teksten og gå dypt inn i den (forstå metateksten). I denne konteksten vurderer de at verktøyet ikke vil gi noen gevinster for deres arbeid. De frykter også at oversettere kan ende opp utelukkende som «etterredigerere». En annen utfordring er at de arbeider med en del tekster som i utgangspunktet er av lav kvalitet. Ofte må de gjøre et grundig forarbeid for å forstå hva budskapet i teksten egentlig er. MT@EC kan gjøre at man i mindre grad vurderer kildeteksten på en kritisk måte.

Videre påpekes det at maskinoversettelse alltid vil kreve at man bearbeider den oversatte teksten og etterredigerer den, og at maskinoversettelsesresultater bør benyttes med forsiktighet.

Samlet sett vurderer vi at behovet de interne oversettere har for MT@EC som begrenset. Bakgrunnen er at de fleste allerede har tatt i bruk profesjonelle oversettelsesverktøy og at MT@EC vil gi dem begrensede gevinster utover det de allerede har oppnådd av effektivitetsgevinster ved å ta i bruk slike verktøy.

3.6.2 Behovet i statlige virksomheter

Vi har også kartlagt behovet hos statlige virksomheter som er de som etterspør oversettelsestjenester. I Tabell 3-4 gjengis noen av innspillene på behovet for MT@EC.

Litt over halvparten av virksomhetene som har besvart spørreskjemaet mener de har behov for MT@EC og at tilgang på MT@EC vil gi virksomheten gevinster. De fleste mener likevel at gevinstene vil være begrensede. Flere virksomheter påpeker at det er vanskelig å besvare spørsmålet uten å vite mer presist hva slags kvalitet man kan oppnå med verktøyet. Virksomhetene er opptatt av at oversettelsen må ha relativt høy kvalitet for at det skal gi de gevinster. Mange påpeker også at de har behov for at maskinoversettelsen er tilpasset terminologien de benytter for at de skal få utbytte av maskinoversettelse. Det bør påpekes at virksomhetenes vurdering trolig preges av hva slags syn de har på maskinoversettelse generelt og hva slags resultater de ser for seg at maskinoversettelse kan gi.

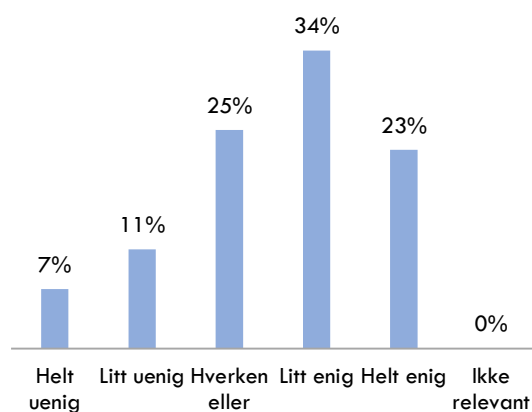
Tabell 3-4: Innspill på behov for MT@EC

Virksomhet	Innspill
NAV	<p>«Med økende fokus på selvbetjeningsløsninger, og økende grad av elektronisk samhandling mellom bruker og NAV, er språklige barrierer en utfordring. Det kan ikke utelukkes at et verktøy for oversettelse som foreslått nedenfor, kan være et bidrag for å avhjelpe de utfordringer som her er nevnt. Men dette må man ev. vurdere grundigere.»</p> <p>«Integrasjon av oversettelser av fagbegreper og juridiske begreper til bruk inn i webløsninger, apper, brev mv. er interessant. NAVs (u)tilgjengelighet på forskjellige språk er en stor kilde til henvendelser i skranke.»</p> <p>«Den skisserte løsningen ser for meg ut å være et stykke unna å dekke behovene for språklig presisjon og teknisk integrasjon (gammelt dokumentfokus).»</p>
Nærings- og fiskeridepartementet	<p>«Det er stor internasjonal interesse for norsk politikk på området [norsk fiskeri- og akvakulturforvaltning, herunder sjømattrygghet], og begrenset kapasitet i NFD og andre aktuelle marine institusjoner til å besvare alle henvendelser. Samtidig er det ønskelig at god norsk forvaltningspraksis på dette området blir delt med flere land. Mye av dokumentasjonen her foreligger bare på norsk, og maskinoversettelse til andre språk kan åpne for selvbetjeningsløsninger der interesserte i andre land kan få tilgang til slik dokumentasjon på sitt eget språk, uten at dette skaper særlig merarbeid i Norge når maskinoversettelsesteknologien er innført og tilstrekkelig kvalitetssikret.»</p>
Lovdata	<p>«Svært lite norsk regelverk er tilgjengelig på annet enn norsk – ikke engang på engelsk. Når regelverk oversettes gjøres det ofte av de store advokatkontorene og er kun tilgjengelige for dem og deres betalende kunder. Stadig flere som bor i Norge har behov for et oversatt regelverk. Også for norsk næringsliv og utenlandsk næringsliv som planlegger aktiviteter i Norge vil et oversatt regelverk ha verdi.»</p> <p>«På [N-Lex] sidene gis informasjon om nasjonal lovgivning og rettssystemet i alle medlemslandene i EU. Det er oversettelser til språkene i alle de 28 medlemslandene som er presentert på siden. Norge i ikke med på N-Lex sidene, men har tilsvarende behov som EU-landene for å presentere seg og sine løsninger. Det bør være et mål for Norge å få innpass på siden eller å presentere seg i ulike språk på egen side.»</p>
Brønnøysundregistrene	<p>«Behovet [for MT@EC] har til nå vært av begrenset karakter, men vi ser at det i framtiden kan bli mer aktuelt for eksempel i forbindelse med IML, epost firmapostmottak, informasjonsbehandling og saksbehandling. Brønnøysundregistrene tror at bruken av MT@EU vil være spredt fra informasjonsutveksling mellom statlige virksomheter i EU både i- og utenfor IML, det å kunne skille ut relevant informasjon fra innkommende dokumenter og videresende dette/disse til rette vedkommende, til bruk som verktøy i internt arbeid og til bruk for dokumenter og presentasjoner.»</p>
Forbrukerombudet	<p>«Det vil kunne være et svært nyttig verktøy i vårt samarbeid med forbrukermyndigheter i andre land. Både i saker som nevnt over, og hvor vi har mer uformell kontakt og orienterer om praksis og pågående prosesser.»</p>
DIBK	<p>«Vi har i liten grad behov for MT@EC. For de fleste i direktoratet brukes hovedsakelig norsk som arbeidsspråk.»</p>

Finanstilsynet	«MT@EC kan senke terskelen for dette [les: oversettelser av enkelte lover, forskrifter og rundskriv] til engelsk, og dermed gjøre det norske lovverket, og Finanstilsynets rundskriv, mer tilgjengelig for omverdenen.»
Datatilsynet	«Vår kommunikasjon med andre lands personvernmyndigheter, bedrifter og enkeltpersoner vil i første rekke foregå på engelsk. Vi vil likevel tro at behovet for denne typen verktøy vil være en del større enn i dag – fordi vi vil kunne måtte ta stilling til tekstmateriale som forelegges oss også på andre språk enn engelsk.»
Finansdepartementet	«Når det gjelder behovet for maskinell oversettelse, er vår vurdering at det må være et høyt presisjonsnivå i oversettelser av de fleste EU-dokumenter, ikke minst av faguttrykk. Det vil være vanskelig å "stole på" at dette blir levert tilfredsstillende gjennom standardiserte programmer. Av den grunn ser vi ikke et stort behov for slikt verktøy.»
UDI	«Om UDI evt. vil benytte et maskinoversettelsesverktøy vil avhenge av kvaliteten på dette systemet og de reelle kostnadene ved innføring og bruk av dette. Dette systemet bør i så fall bruke verktøy som er utviklet for å sikre at det i ulike språk brukes riktige betegnelser på begreper/fenomener/institusjoner.»
Skattedirektoratet	«Vi har komplekse regelverk med mange faguttrykk. [...] Ut fra de forutsetningene vi har for å svare på dette, så vil vi si nei til at vi har behov for et maskinoversettelsesverktøy. Dog er det mulig at maskinoversettelsesverktøy i kombinasjon med en database med faguttrykk med tilhørende oversettelser vil være nyttig for informasjon som er mindre sammensatt enn mange av regelverkstekstene som finnes i dag.»

Vi har også i spørreundersøkelsen blant ansatte kartlagt behovet for maskinoversettelse.

Figur 3-11: Fordeling utvalg: «Hvor enig er du i påstanden? «Jeg har behov for et maskinoversettelsesverktøy i mitt arbeid utover de hjelpemidlene jeg har tilgang på i dag.»), 44 svar



Kilde: Spørreundersøkelse blant ansatte i statlige virksomheter

57 prosent er litt enig i eller helt enig i påstanden om at de har behov for maskinoversettelse utover de hjelpemidlene de har tilgang på i dag.

Respondentene i spørreundersøkelsen ble også spurt om hvorvidt de er enig i påstanden: «Det er sannsynlig at jeg ville benyttet et maskinoversettelsesverktøy for

å oversette tekster mellom norsk og [andre EU-språk enn engelsk] / [engelsk], selv om maskinoversettelsen ikke er perfekt.» Henholdsvis 70 prosent sa seg helt eller delvis enig i at de ville bruke verktøyet til oversettelser fra norsk bokmål til engelsk, mens ca. 50 prosent vil bruke verktøyet til å oversette fra norsk bokmål til andre språk enn engelsk. Videre ble respondentene spurt om de vil bruke verktøyet til å forstå tekster på andre EU-språk enn engelsk. Rundt 65 prosent av respondentene sa seg helt eller delvis enig i dette. Hovedparten av respondentene i spørreundersøkelsen ser dermed på maskinoversettelse som nyttig.

Generelt synes det å være et stort behov for å oversette tekster ute i virksomhetene. Tilbakemeldingene tyder også på at behovet for oversettelser er større enn det som oversettes i dag. Det tyder på at det er behov for å effektivisere prosessen med å oversette tekster.

Samlet sett er behovet for en løsning for automatisk oversettelse i mange virksomheter stort, men det er behov for relativt høy kvalitet på oversettelsen og at verktøyet er tilpasset det aktuelle domenet.

4. Alternativer for en nasjonal satsing på norsk i MT@EC

4.1 Språk og fagområder

MT@EC er bygget opp slik at hvert språkpar har én database. Det er imidlertid ikke slik at enhver språkkombinasjon har fått opprettet egne oversettelsesmotorer og tilhørende minne. Som hovedregel forgår alle oversettelser via engelsk. Dersom man eksempelvis ønsker å oversette en tekst fra norsk bokmål til spansk, vil MT@EC først oversette fra norsk bokmål til engelsk og deretter fra engelsk til spansk.

Dersom alle språkpar skulle hatt direkte oversettelsesmuligheter, vil dette innebære omtrent 600 oversettelsesmotorer. Til tross for at opprettelsen av den tekniske oversettelsesmotoren ikke er krevende, gitt at et tilstrekkelig stort corpus finnes, medfører den direkte språkforbindelsen store vedlikeholdskostnader. Grunnet ressursbegrensninger påpeker EU-kommisjonen¹⁰ to hovedkriterier for å opprette en direkte språkforbindelse: omfanget av oversettelser mellom de to språkene og størrelsen på tilgjengelig corpora.

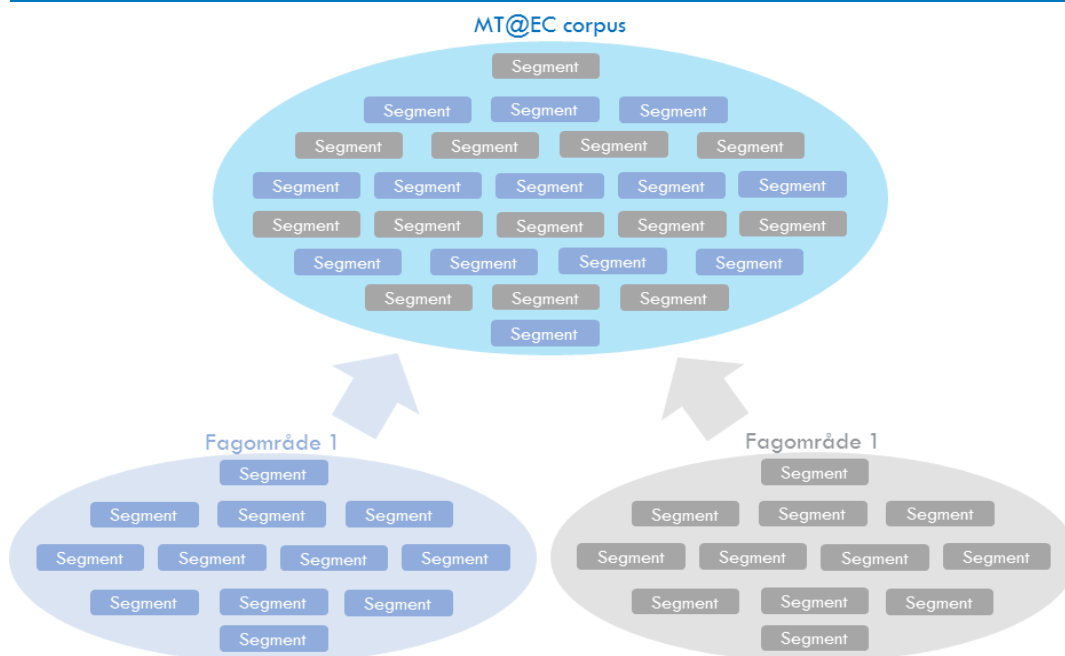
Innspillene vi har fått fra EU-kommisjonen tilsier at det er lite sannsynlig at de vil prioritere å bygge

direktemotorer mellom norsk bokmål og et annet språk ettersom oversettelser til/fra norsk bokmål og andre språk enn engelsk, i et europeisk perspektiv, har et svært lite omfang. Dette gjelder uavhengig av om tilstrekkelig corpora på norsk bokmål og andre EU-språk etableres. Vi avgrensner derfor satsingsalternativene til språkkombinasjonen norsk bokmål-engelsk.

Ettersom nynorsk er et offisielt språk i Norge, vil det være mulig å opprette en direkte oversettelsesmotor mellom nynorsk og engelsk uavhengig av omfanget til oversettelser til/fra nynorsk. Det er imidlertid fortsatt en forutsetning at det finnes tilstrekkelig corpora som kan benyttes, der EU-kommisjonen foreløpig krever en størrelse på minst 100 000 segmenter. I tillegg til satsingen på bokmål vil vi derfor også vurdere en eventuell satsing på nynorsk.

MT@ECs minne kan eksempelvis bestå av corpora fra ulike fagområder. I minnet til MT@EC samles alle segmenter. I corpora på norsk bokmål-engelsk finnes for øyeblikket kun oversettelsene av EØS-regelverket. Det er mulig å utvide ved eksempelvis å inkludere fagområdet samferdsel ved å legge til corpuset til Statens vegvesen. Dette er illustrert i Figur 4-1.

Figur 4-1 Corpora og segmenter i MT@EC



Kilde: Oslo Economics

¹⁰ Ved Kimmo Rossi i DG Connect

Ettersom MT@EC er statistikkbasert, er antallet segmenter i første omgang den avgjørende faktoren for kvaliteten til maskinoversettelsen.

For at oversettelsen på et gitt fagområde skal være ha god kvalitet, krever det også at terminologien på et gitt fagområde er godt nok representert i corpus. I den videre satsingen vil EU-kommisjonen som nevnt bygge oversettelsesmotorer som er skreddersydd ulike fagområder/domener. I satsingsforslagene vi utarbeider, skiller vi ikke konkret på hvilke fagområder man samler inn tekster fra. Skillet vil snarere gå på i hvilken grad man har mulighet til å produsere corpora på fagområder hvor behovet er stort.

Corpora på norsk bokmål-engelsk består av over 700.000 segmenter og det finnes derfor ingen nedre begrensning på størrelsen av individuelle corpora som kan tas imot ved en eventuell avlevering fra offentlige virksomheter. For at det skal være hensiktsmessig for EU-kommisjonen å benytte ressurser til innarbeiding av de nye segmentene, burde avleveringen av segmenter være av en viss størrelse (trolig over 5000 til 10 000 segmenter).¹¹

4.2 Satsingsalternativene

En satsing på MT@EC vil ikke være et «enten eller»-valg, men heller en vurdering av *i hvilken grad* man burde satse. For å tydeliggjøre mulighetsrommet for en eventuell norsk bokmål satsing på MT@EC, har vi delt analysen inn i tre ulike satsingsalternativer, som beskrevet i Tabell 4-1. Alternativene er ulike nivåer på satsingen som vurderes opp mot hverandre, og

Tabell 4-1 Alternativer for satsing

Satsingsnivå 0	Satsingsnivå 1	Satsingsnivå 2	Satsingsnivå 3	Nynorsksatsing
Ingen endring i antall segmenter	Kontinuerlig avlevering av nye (fremtidige) segmenter etableres	Kontinuerlig avlevering av nye (fremtidige) segmenter etableres	Kontinuerlig avlevering av nye (fremtidige) segmenter etableres	Kontinuerlig avlevering av nye (fremtidige) segmenter etableres
	Innhenting av eksisterende filer i rett filformat	Innhenting av eksisterende filer i rett filformat	Innhenting av eksisterende filer i rett filformat	
		Innhenting av parallelle tekster i andre filformat	Innhenting av parallelle tekster i andre filformat	
			Produksjon av corpora	Produksjon av corpora

Kilde: Oslo Economics

¹¹ Forbruker Europa informerte oss om at de hadde produsert 10.000 segmenter innen forbrukerfeltet, men det

opp mot å fortsette med dagens nivå. Det siste er det som kalles «nullalternativet» i analysen. I dette avsnittet vil vi skissere innholdet i alternativene, men kostnadsestimater er gitt i kost-/nytteanalysen i kapittel 5.

I nullalternativet legger vi til grunn at det ikke satses på en ytterligere innsamling av corpora, utover EØS-regelverket som allerede er innhentet. I alle andre satsingsnivå legger vi til grunn at det etableres en kontinuerlig avlevering av segmenter som produseres i fremtiden.

I satsingsnivå 1 forutsetter vi at eksisterende segmenter i rett teknisk filformat innhentes fra oversettere i både offentlige og private virksomheter. Denne segmentinnsamlingen beskrives nærmere i avsnitt 4.2.2, og vi legger til grunn at den inngår i samtlige satsingsalternativ.

Som et neste trinn i en satsing på MT@EC, i form av satsingsnivå 2, vil det være hensiktsmessig å innhente oversatte tekster også i andre tekniske filformat enn de som benyttes av MT@EC. Hva en slik innhenting innebærer er nærmere beskrevet i 4.2.3.

En siste opptrapping av satsingen til nivå 3 vil dermed innebære at det i tillegg produseres corpora, kun med formål om å anvendes i MT@EC. Ved å produsere corpora vil man ha mulighet til å utvide corpora på fagområder som i liten grad er representert i tekstene som samles inn i satsingsnivå 1 og 2. På denne måten kan man utvide anvendelsesområdene til verktøyet. Dette beskriver vi nærmere avsnitt 4.2.4.

Vi vurderer også en mulig satsing på nynorsk.

er uvisst om dette er innarbeidet i MT@EC på nåværende tidspunkt.

Alle satsingsnivåer innebærer at en utpekt instans vil ta imot og samle inn corpora fra offentlige virksomheter. Språkbanken er i denne sammenheng kanskje spesielt aktuell, ettersom de tilbyr digitale språkressurser til forskning og språkteknologisk utvikling.¹² Språkbanken, som er underlagt Nasjonalbiblioteket, var også ansvarlig for innsamlingen og avleveringen av Norges eksisterende corpus.

4.2.1 Kontinuerlig avlevering av nye segmenter

Med kontinuerlig avlevering menes en avlevering av segmenter som produseres i fremtiden. En slik innsamling vil dermed ikke kreve en full gjennomgang av eksisterende minner, eksempelvis for å fjerne personsensitive opplysninger. Videre menes kun en avlevering av segmenter i rett teknisk format, som i praksis innebærer at kun profesjonelle oversettere (som benytter profesjonelle oversettelsesverktøy i sitt arbeid) blir berørt.

I denne avleveringsprosessen vil virksomhetene kontinuerlig kunne samle opp filer til overlevering i sitt arbeid. Dette innebærer i praksis kun å lagre kopier av de relevante filene på ett bestemt sted. Deretter vil en utpekt instans kunne samle inn disse filene regelmessig, eksempelvis én gang i året. Oversetterne vil fortløpende kunne vurdere hvilke filer som inneholder sensitive opplysninger og hvilke filer som kan overleveres, slik at ingen gjennomgang av segmentminner er nødvendig ved overlevering.

Selv om en virksomhet ikke oversetter selv, men benytter eksterne oversettere, vil det være mulig å innhente segmenter fra virksomheten. Eksempelvis kan kontrakten for oversettelsesoppdraget eksplisitt inneholde en klausul om at eierskapet til oversettelsen, inkludert det tekniske filformatet, tilhører virksomheten. Ved hvert oppdrag vil dermed også eksempelvis TMX-filer være en del av leveransen fra oversettelsesbyrået.

UDs enhet for EØS-oversettelser opplyser om at eierskap også over segmentminnene allerede er inkludert i deres kontrakter med eksterne leverandører. Dermed fremstår det som lite ressurskrevende å skulle inkludere en tilsvarende tekst i standardkontrakter for oversettelsestjenester ved alle offentlige virksomheter.

Etttersom en avlevering av fremtidige segmenter vil være en lite ressurskrevende satsing, legger vi den til grunn i alle satsingsalternativ. Satsingen innebærer ingen gjennomgang av eksisterende minner, men kun kopiering av fremtidige oversettelsesminner.

4.2.2 Innhenting av eksisterende filer i rett filformat

Segmentminnene fra profesjonelle oversettere, som har benyttet et profesjonelt oversettelsesverktøy, har fordelen at de kan benyttes direkte i MT@ECs database. Filene må dermed ikke konverteres eller redigeres i forbindelse med innsamlingen. Videre er disse filene enkelt å finne, ettersom de allerede ligger i databasen til oversettelsesverktøyene. Dermed er det ikke nødvendig å benytte ressurser for å lete etter oversatte dokumenter eller filer.

Profesjonelle oversettere i både offentlige og private virksomheter besitter segmenter. Dermed finnes allerede segmenter fra en rekke ulike fagområder i dag, blant annet innen samferdsel (Statens vegvesen), sjøfart (Sjøfartsdirektoratet), utdanning (Universitetet i Oslo) og juss (seksjon for EØS-oversettelser i Utenriksdepartementet). I tillegg har BLD produsert segmenter innen forbrukerspørsmål, som inneholder mer dagligdags språk. Ved en innsamling av eksisterende segmenter i rett filformat vil dermed språk tilpasset en rekke nye fagområder inkluderes i MT@EC.

Innhenting av corpora fra offentlige virksomheter

Innhenting av filer i rett filformat innebærer at profesjonelle oversettere i offentlige virksomheter foretar en gjennomgang av sine segmentminner. Dette er fordi det kun er profesjonelle oversettere som benytter profesjonelle oversettelsesverktøy, og som dermed har segmenter i rett filformat.

Ved en innsamling av segmentminner fra offentlige virksomheter vil det være nødvendig med en gjennomgang av filene. Ved en slik gjennomgang vil eventuelle personopplysninger, sikkerhetsgradert informasjon og liknende fjernes fra minnene før disse kan overleveres.

Innhenting av corpora fra private oversettere

Private oversettere vil også besitte segmentminner fra sine offentlige oppdrag. Så langt det er mulig, vil det dermed være ønskelig å innhente segmenter også fra de private aktørene.

I henhold til Statens standardavtaler om konsulenttjenester SSA-O, tilfaller all eiendomsrett til oppdragets resultater kunden. Dersom eiendomsretten også omfatter minnet, innebærer det at den statlige enheten kan kreve utlevert sitt minne ved bytte av eksternt leverandør. Amesto Translations opplever som privat leverandør av oversettelsestjenester at dette sjeldent eller aldri forekommer. Kontrakter på eksternt oversettelse settes ut på anbud minimum hvert 4. år, og et mer bevisst forhold til hva staten eier, hadde bidratt til større besparelser.

¹² <http://www.nb.no/Tilbud/Forske/Spraakbanken>

Juridisk eierskap til slike segmentminner fremstår i praksis som uvisst, i de tilfeller der eierskap til de tekniske filene ikke er særskilt avklart i kontrakten for oversettelsesoppdraget. Det er derfor usikkert i hvilken grad oversettelsesbyråene i realiteten vil overlevere segment-minner, og en juridisk avklaring på dette punktet vil være nødvendig.

4.2.3 Innhenting av parallelle tekster i andre filformat

Det er også mulig å innhente andre oversettelser, som ikke er utført av profesjonelle oversettere med profesjonelle oversettelsesprogrammer. Eksempelvis kan også Word-, Excel- og PowerPoint-filer samles inn. I tillegg vil det være mulig å benytte menyer og tekster fra nettsider som er tilgjengelige på bokmål og engelsk, eksempelvis fra SSB og Norge.no. For at disse filene skal kunne inngå i MT@ECs database må parallelle filer foreligge. Med parallelle filer menes at det samme dokumentet foreligger i to versjoner – ett på norsk og ett på engelsk.

Innhentede parallelle dokumenter må deretter konverteres til rett filformat. Dette kan eksempelvis gjøres ved at tekstene benyttes i et profesjonelt oversettelsesverktøy som SDL Trados Studio. I oversettelsesverktøyet må hver setning, overskrift og hvert kulepunkt på norsk manuelt matches med tilsvarende engelske oversettelse. Dette konverteringsarbeidet krever dermed ressursbruk, som ikke er nødvendig ved innhentede filer som allerede har rett filformat.

Ettersom virksomheter som bestiller eksterne oversettere ikke alltid har bruk for tekstversjonen på originalspråket, er det usikkert om originalfilene er tilgjengelige. En oversettelse alene har ingen verdi uten at den parallelle teksten på originalspråket foreligger. Trolig vil ikke disse dokumentene være samlet på ett bestemt sted slik tilfellet er for segmentfilene i profesjonelle oversettelsesverktøy. Det vil derfor være relativt tidkrevende å fremskaffe og bearbeide oversatte tekster på andre tekniske filformat.

Ettersom det er knyttet stor usikkerhet til omfanget og tilgjengeligheten av parallelle tekster, er det også knyttet stor usikkerhet til antallet segmenter som kan innhentes. Fordelen er at innhentede segmenter vil kunne ha stor faglig bredde med tanke på ordforråd og fagområder.

4.2.4 Produksjon av corpora

Dersom innhentede corpora fra satsingsnivå 1 og 2 ikke inneholder tilstrekkelig stort ordforråd, kan tekster oversettes for å inngå som segmenter i MT@ECs minne. Dette gjør at man kan utvide anvendelsesområdene til MT@EC ved å få

tilstrekkelig corpora på områder som ikke er godt nok dekket av det som samles inn i satsingsnivå 1 og 2.

Ved produksjon av corpora vil det være mest hensiktsmessig å fokusere på fagområder hvor ordforrådet i liten grad dekkes av eksisterende corpora. Som drøftet i behovskapittelet er det en stor bredde i type tekster som oversettes av virksomhetene og også stor spredning i hvilken fagterminologi det er behov for.

Ettersom satsingsnivå 1 og 2 innebærer en innsamling av segmenter fra en rekke ulike områder inkludert EØS-regelverket, kan det forventes at ordforrådet på en rekke fagområder blir tilstrekkelig dekket. Dette inkluderer samtlige fagfelt som er inkludert i EØS-regelverket i tillegg til fagområdene samferdsel (gjennom Statens vegvesen og Jernbaneverket), sjøfart (gjennom Sjøfartsdirektoratet), utdanning (gjennom Universitetet i Oslo), helse og omsorg (gjennom Helse- og omsorgsdepartementet), samt næringssektoren (gjennom Nærings- og fiskeridepartementet).

I tillegg til ordforrådet innen de ovennevnte fagområdene, vil også terminologien innen jus, administrasjon og økonomi få god dekning. Vi vil derfor ikke anbefale å produsere corpora på disse fagområdene i første omgang.

Hverdagsspråk vil derimot i liten grad være dekket i corpora som samles inn i satsingsnivå 1 og 2 ettersom disse tekstene i stor grad er offisielle dokumenter hvor slikt språk ikke anvendes. Ved anvendelse av MT@EC i ulike webtjenester som eksempelvis i ODR er tanken at MT@EC blant annet skal benyttes til å oversette tekst som skrives inn i tekstfelter for kommunikasjon mellom ulike brukere. Slik anvendelse krever at hverdagsspråk er godt nok dekket i corpora. Vi anbefaler derfor at produksjon av corpora i første omgang fokuserer på å øke omfanget av hverdagsspråk.

Eksempelvis vil det være mulig å oversette et utvalg av forbrukerklager eller andre tekster med hverdagsspråk. Som tidligere beskrevet er det allerede oversatt omtrent 10 000 segmenter innen forbrukerfeltet. EU-kommisjonen anslår at det kreves omtrent 100 000 segmenter for å sikre tilstrekkelig oversettelses kvalitet. I satsingsnivå 3 legger derfor inn at det produseres 100 000 segmenter med hverdagslig språk innen forbrukerfagfeltet og andre relevante områder.

4.2.5 Nynorsksatsing

Som følge av at Norge er et medlemsland i EØS og nynorsk er et offisielt språk i Norge, er det mulig å få en oversettelsesmotor for nynorsk, på lik linje med bokmål. Ettersom det i dag kun finnes omtrent 50 000 segmenter mellom engelsk og nynorsk, kreves dermed

ytterligere 700 000 segmenter for å få samme omfang på corpora som for bokmål.

Satsingsalternativet legger til grunn at nynorsk likestilles med nullalternativet i bokmålsatsingen. Dersom en egen oversettelsesmotor for nynorsk etableres, vil det antagelig også være hensiktsmessig at nye oversettelser til/fra nynorsk avleveres kontinuerlig.

Som drøftet i kapitlet om behov oversetter virksomhetene i liten grad til/fra nynorsk i dag. Omfanget av eksisterende oversettelser mellom nynorsk og EU-språk som kan benyttes som corpora kan antas å være svært begrenset basert på våre undersøkelser. For at segmentminnet for nynorsk skal være på samme størrelse som dagens segmentminne for bokmål, vil det derfor kreve at corpora produseres. Ved produksjon av corpora kan relevante fagområder velges ut, for å sikre av corpuset får tilstrekkelig ordforråd og faglig bredde.

5. Kost-/nytteanalyse av ulike satsingsnivå på MT@EC

5.1 Identifiserte nytte- og kostnadsvirkninger av en satsing på MT@EC

I det følgende presenterer vi hvilke nytte- og kostnadsvirkninger vi forventer vil oppstå ved en satsing på MT@EC. Vi skiller her på prissatte og ikke-prissatte virkninger. Det er primært kostnadseffektene vi vurderer at det er hensiktsmessig og forsvarlig å tallfeste i form av prissetting. De øvrige virkningene av en satsing behandles som ikke-prissatte virkninger.

5.1.1 Ikke-prissatte virkninger

I det følgende gir vi en oversikt over de ikke-prissatte nytte- og kostnadsvirkningene en satsing på MT@EC forventes å gi.

Mer kommunikasjon og informasjonsdeling på tvers av landegrensene

MT@EC vil gjøre det mindre tidkrevende å oversette og forstå tekster. De statlige virksomhetene melder om at behovet for å oversette tekster er større enn omfanget av oversettelser som faktisk blir gjennomført på grunn av ressursbegrensninger.

Tilgang på MT@EC forventes å bidra til at flere tekster oversettes til/fra norsk (bokmål eller nynorsk), samt til å forstå dokumenter på andre språk. Dette kan videre bidra til at norsk statlig sektor i større grad enn i dag vil dele flere tekster og dokumenter med andre land, samt i større grad enn i dag benytte tekster og dokumenter på andre språk i det daglige arbeidet.

Jo flere virksomheter innenfor ulike fagområder som avleverer segmenter, jo flere anvendelsesområder vil MT@EC kunne få. Produksjon av corpora vil gi gevinster i form av at MT@EC vil kunne anvendes på områder hvor det ikke foreligger oversatte tekster i stort nok omfang til at man kan sikre tilstrekkelig kvalitet på maskinoversettelsene. Dette gjør at man kan skape mer kommunikasjon og informasjonsdeling på stadig flere områder.

Ved å benytte MT@EC i EUs digitale tjenester og plattformer vil det bidra til tilrettelegge for enklere og dermed trolig mer kommunikasjon og informasjonsdeling på de ulike områdene tjenestene gjelder. Dette vil gi opphav til gevinster for de ulike områdene tjenestene dekker. Eksempelvis vil ODR bidra til enklere kommunikasjon mellom ulike parter i behandlingen av forbrukersaker på tvers av land. Dette vil videre kunne gi opphav til at forbrukerrettigheter ivaretas på en bedre måte ved at flere velger å klage og/eller ved at det blir mindre tidkrevende å klage.

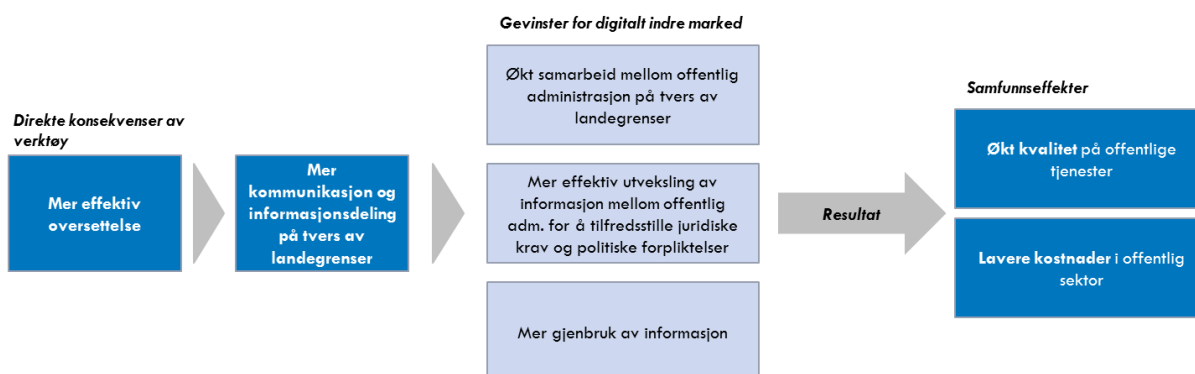
Mer kommunikasjon og informasjonsdeling på tvers av land vil med andre ord gi opphav til positive samfunns effekter på flere områder. Vi har illustrert de positive samfunns effektene dette kan generere i Figur 5-1. For det første kan det bidra til mer samarbeid mellom offentlig administrasjon i ulike land. I mange politiske og juridiske prosesser kreves det at informasjon utveksles mellom ulike parter. Eksempelvis i forbrukersaker på tvers av land. Ved maskinoversettelse kan informasjonen oversettes og deles mer effektivt mellom partene. Mer gjenbruk av informasjon på tvers av land vil også gi opphav til gevinster. Eksempelvis kan en etat i Norge som skal utvikle en ny veileder, ta utgangspunkt i arbeid som er gjort i andre land.

Samlet sett forventes maskinoversettelse å gi opphav til økt kvalitet på offentlige tjenester og lavere kostnader i offentlig sektor.

Mer standardisering av språk

MT@EC vil kunne ha gevinster for oversettelsene til de profesjonelle oversetterne i staten ved at oversettelsene i større grad bygger på et felles corpus. Ettersom oversettelsen basert på MT@EC bygger på tidligere oversettelser, vil bruk av verktøyet innebære at man blir mer konsekvent i ordbruken, og dette bidrar dermed til standardisering av språk i offentlige dokumenter. Dette vil videre gi kvalitetsgevinster på språkområdet som kan ha en verdi i seg selv, eksempelvis ved tydeligere budskapsformidling fra offentlige virksomheter.

Figur 5-1: Samfunnseffekter av flere oversettelser



Kilde: Oslo Economics

Denne virkningen vil bli forsterket av at MT@EC tas i bruk på nye områder. Ved at flere virksomheter avleverer sine segmenter vil både ordforrådet og størrelsen på corpuset øke. Dermed vil MT@EC kunne anvendes på nye fagområder, der ordforrådet før satsingen ikke var stort nok til å sikre tilstrekkelig kvalitet på maskinoversettelsene, og flere virksomheter vil bli delaktige i standardiseringen av språket.

Lavere språklig kvalitet

På den andre siden kan MT@EC også ha negative effekter på språklig kvalitet ved at ordene som benyttes er feil eller uegnet eller ved at det er grammatiske feil i oversettelsen. MT@EC kan også bidra til at språket blir mindre variert ettersom maskinoversettelsen per definisjon velger det mest vanlige ordet. Dette kan på sikt ha negative konsekvenser for den språklige kvaliteten på tekster i offentlig forvaltning.

Flere misforståelser og mer støy i kommunikasjonen på tvers av landegrensar

Maskinoversettelse kan bidra til å skape støy ved at oversettelsen blir feil i konteksten, man ikke benytter riktig fagterminologi eller mangler viktige ord. Ved maskinoversettelse går man heller ikke inn og analyserer den opprinnelige teksten nærmere. Dersom kvaliteten på teksten i utgangspunktet er lav, kan maskinoversettelse gjøre budskapet enda mer uklart. Ved ukritisk bruk eller mangelfull etterredigering kan MT@EC dermed bidra til feilkommunikasjon, misforståelser og støy mellom ulike parter som potensielt kan ha betydelige konsekvenser.

Bedre ivaretagelse av personvern og samfunnssikkerhet

Som kartlagt i behovskapittelet er gratis oversettelsesverktøy som Google Translate og Bing Translate relativt utbredt i statlige virksomheter.

I Google Translate's brukervilkår står det blant annet:¹³

- “When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content.”
- “Some of our Services allow you to upload, submit, store, send or receive content. You retain ownership of any intellectual property rights that you hold in that content. “

Bruken av disse verktøyene i offentlig sektor kan potensielt være problematisk. For det første fordi at sensitive opplysninger kan komme på avveie. For det andre på grunn av at man gir Google og evt. partnere rettigheter til å benytte innholdet, ref. utdragene fra brukervilkårene over.

Vi har ikke grunnlag for å si noe om omfanget på problematisk bruk av disse verktøyene. Dersom uegnet bruk av gratis maskinoversettelse er utbredt kan det å gi virksomhetene tilgang på maskinoversettelse gjøre at man i større grad enn i dag ivaretar personvern og samfunnssikkerhet. Størrelsen på gevinsten avhenger av hvor stort problemet med bruk av Google Translate og Bing Translate er opprinnelig.

¹³ <https://www.google.com/policies/terms/>

Tilgjengeliggjøring av informasjon og dokumenter på nynorsk

Ved en nynorsksatsing i MT@EC kan verktøyet benyttes til å oversette tekst mellom nynorsk og EU-språk. Dette vil ha en verdi for samfunnet ved at mer tekst blir tilgjengeliggjort på nynorsk. Målloven har som krav at 25 prosent av tekstene i offentlig forvaltning skal være på nynorsk.

Det finnes allerede verktøy som oversetter til nynorsk som eksempelvis Nyno så dermed kan allerede potensialet for tilgjengeliggjøring av informasjon på nynorsk ved maskinoversettelse være tatt ut.

Konkurranse i markedet for maskinoversettelse

En satsing på EU-kommisjonens verktøy for maskinoversettelse vil også kunne ha konsekvenser for eventuelle private aktører som tilbyr maskinoversettelsesverktøy i markedet.

Konkurranse er et virkemiddel for at samfunnet kontinuerlig skal oppnå forbedringer i hvordan oppgaver løses, samtidig som at de ressursene som kreves for å løse oppgavene blir brukt på best måte. Dersom en satsing på MT@EC hemmer konkurransen i markedet for maskinoversettelse kan det resultere i dyrere og/eller lavere kvalitet på løsningene for maskinoversettelse som tilbys i markedet.

Dersom det offentlige tilbyr en tjeneste under markedspris kan det skape en vridning i etterspørselen fra det private tilbudet til det offentlige tilbudet. I dette tilfellet tilbyr EU-kommisjonen MT@EC gratis til offentlig forvaltning i land som bidrar med corpora.

Det finnes en rekke ulike private tilbydere av løsninger for maskinoversettelse. I tillegg til Google Translate og Bing Translate finnes det ulike andre private leverandører av maskinoversettelse. Eksempelvis er Apertium et regelbasert maskinoversettelsesprogram, som oversetter mellom ulike EU-språk. Foreløpig oversetter verktøyet kun mellom bokmål/nynorsk og dansk. Bruksområdene for CAT-verktøyene, som benyttes av de profesjonelle oversetterne, vil også delvis overlape med bruksområdet til MT@EC. Begge verktøyene er løsninger for automatisk oversettelse.

Ettersom det finnes private tilbydere som tilbyr løsninger som i større eller mindre grad er substitutter til MT@EC kan en satsing på MT@EC potensielt skape konkurransevridende effekter.

Hvor sterk den negative virkningen av konkurransevridning faktisk vil bli, avhenger primært av hvor nære substitutter MT@EC og øvrige maskinoversettelsesverktøy er. Dersom tilbudet fra private aktører i liten grad er et substitutt til EU-kommisjonens maskinoversettelse vil den konkurransevridende effekten bli mindre.

Nedenfor belyser vi noen momenter som er sentrale i vurderingen av hvor nære substitutter MT@EC er til private løsninger. Vi vil understreke at dette er en overordnet vurdering.

For det første har offentlig forvaltning behov for en sikker løsning av høy kvalitet, som kan benyttes for oversettelse av tekst med konfidensielle opplysninger. Google Translate vil eksempelvis være lite egnet til oversettelse av dokumenter med sensitive opplysninger. Det kan imidlertid tenkes at andre private aktører vil være i stand til å skape løsninger som ivaretar behovet for sikkerhet.

Offentlig forvaltning har behov for at oversettelsene benytter en terminologi som er tilpasset språket i offentlig forvaltning. EU-kommisjonen peker her på at MT@EC i større grad enn andre alternativer i markedet kan benytte en terminologi som er korrekt i ulike settinger. Dersom øvrige aktører fikk tilgang på språkressursene MT@EC benytter, kan det imidlertid tenkes at også de kunne utviklet oversettelsesprodukter som er domenespesifikke.

En nærmere utredning av de konkurransemessige effektene ved MT@EC er utenfor mandatet til denne rapporten. Vi vurderer at de potensielle konkurransevridende effektene MT@EC gir, avhenger av hvordan satsingen utformes. Det finnes måter man kan redusere de konkurransevridende effektene på. Eksempelvis ved å gi øvrige aktører tilgang til de samme språkressursene som MT@EC benytter. Forskningsmiljøene påpeker at tilgang på språkressurser er viktig for at forskningsmiljø og andre skal kunne forbedre og utvikle nye løsninger for maskinoversettelse.

5.1.2 Prissatte virkninger

I Tabell 5-1 gir vi en oversikt over de prissatte virkningene en satsing på MT@EC forventes å gi. Vi har vurdert at det kun er hensiktsmessig med en prissetting av kostnadene knyttet til en satsing på MT@EC. Vi redegjør for forutsetningene vi tar ved prissetting av virkningene i kapittel 0.

Tabell 5-1: Prissatte virkninger ved en satsing på MT@EC

Virkninger av maskinoversettelse	Beskrivelse
Kostnader knyttet til årlig avlevering av segmenter	Interessentene anslår det som lite tidkrevende å legge til rette for kontinuerlig fremtidig avlevering av segmenter. En slik tilrettelegging vil innebære at de profesjonelle oversetterne fortløpende kopierer minnefiler som kan overleveres, og siler ut segmenter og/eller filer som inneholder personopplysninger eller sensitiv informasjon. I tillegg til utgifter i virksomhetene vil det også oppstå driftskostnader sentralt for etaten som blir ansvarlig for innsamlingen.
Kostnader knyttet til avlevering av eksisterende segmenter	Sammenliknet med kontinuerlig avlevering av segmenter som produseres i fremtiden, anser interessentene våre det som mer krevende å foreta en gjennomgang av eksisterende minner for å fjerne personopplysninger, forretningshemmeligheter fra anbuds konkurranser eller andre sensitive opplysninger. Noen virksomheter har organisert segmentfilene i underkategorier, der noen kategorier er mer uproblematisk enn andre. Videre finnes få virksomheter med interne profesjonelle oversettere, og dermed blir det samlede ressursbehovet for gjennomgang av segmentfiler begrenset.
Kostnader knyttet til avlevering og konvertering av oversatte tekster på andre filformat	Informasjon innhentet gjennom dybdeintervjuene tyder på at det vil være svært tidkrevende å finne parallelle tekster i andre filformat. Årsaken til dette vil være at slike tekster i et samlet på ett sted og at det kan være utfordrende å finne både originaldokumentet i tillegg til oversettelsen. Videre må tid til konvertering påberegnes.
Kostnader ved produksjon av corpora	For produksjonen av corpora legger vi til grunn en gjennomsnittlig markedspris på 2 kr per ord og 9 ord per segment.

5.2 Sammenhengen mellom corpora-størrelse og nytteverdi

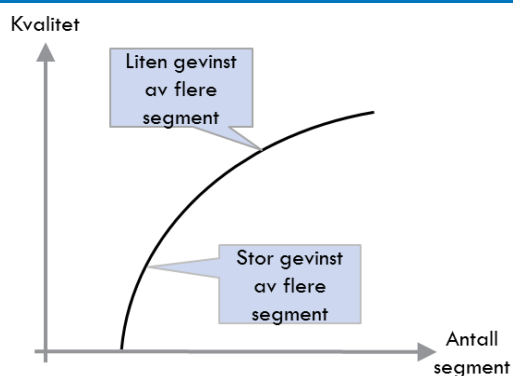
Ettersom MT@EC er statistikkbasert, vil verktøyet virke bedre jo større corpuset er. Videre vil nytteeffekten fra nye segmenter være mindre, jo større corpus som finnes fra før av. Eksempelvis vil Norge, med et relativt lite corpus på omtrent 750.000 segmenter, oppleve en stor forbedring i kvaliteten på maskinoversettelsen dersom corpuset utvides. Et land som derimot allerede har et stort corpus, som Sverige med sine 40 millioner segmenter, vil derimot oppleve at nye segmenter ikke fører til store forbedringer.

Antallet segmenter er ikke den eneste faktoren som påvirker nytteverdien til maskinoversettelsen. Et bredt utvalg av tekster fra ulike fagområder og med ulikt ordforråd vil også ha betydning for maskinoversettelsens kvalitet.

I alle satsingsnivå legges det til grunn at segmenter innhentes fra en rekke ulike offentlige virksomheter. Dermed vil alle satsingsnivå innebære at det innhentes corpora fra en rekke ulike fagområder. Eksempelvis har BLD produsert segmenter for å forbedre MT@ECs datagrunnlag knyttet til forbrukerklager i webportalen ODR (online dispute resolution).

Ved en innhenting av segmenter i rett filformat, som er en del av alle satsingsnivå, vil dermed segmenter innen forbrukerklager inngå, i tillegg til segmenter fra sjøfart, samferdsel og juss. Ettersom alle satsingsalternativ inkluderer segmenter fra en rekke ulike fagområder, vurderer vi antall segmenter som en god indikator for nytteverdien av satsingen. Sammenhengen mellom segmenter og nytte er illustrert i Figur 5-2.

Figur 5-2 Sammenheng mellom segmenter og nytte



Kilde: Oslo Economics

5.2.1 Estimert antall segmenter

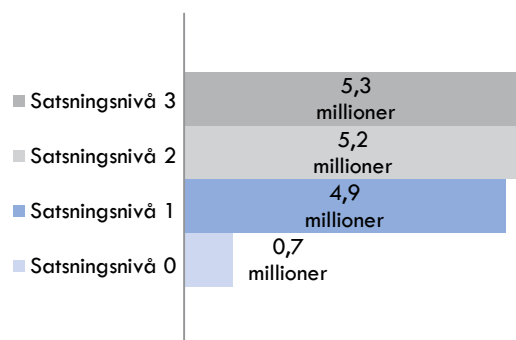
Figur 5-3 gir en oversikt over estimert antall segmenter i et 10-års perspektiv ved de ulike satsingsnivåene. Figuren illustrerer at det oppstår en stor økning i antall segmenter ved å gå fra nullalternativet til satsingsnivå 1. En ytterligere økning utover satsingsnivå 1 vil derimot gi liten gevinst i form av flere segmenter.

I det følgende redegjør vi for estimeringen av antall segmenter i de ulike satsingsalternativene. En samlet oversikt over tallene som ligger til grunn for Figur 5-3 er gitt i Tabell 5-2.

I nullalternativet ligger dagens corpus til grunn. Dette corpuset inneholder, ifølge EU-kommisjonen, i dag 733 081 segmenter. Dette corpuset er inkludert i alle satsingsalternativene. I tillegg har vi inkludert en årlig levering av 300 000 segmenter. Tallet er estimert med bakgrunn i at Amesto i fjor leverte omtrent 2,7 millioner ord til offentlige kunder. Dersom vi forutsetter 9 ord per segment¹⁴, innebærer dette en produksjon av rundt 300 000 segmenter. Grunnet personvern hensyn og liknende vil ikke alle disse segmentene kunne avleveres. Derimot produserer både interne profesjonelle oversettere og andre private oversettelsesselskap også segmenter som kan avleveres. Dermed anser vi 300 000 avleverte segmenter årlig som et minimumsestimat.

¹⁴ Norske setninger inneholder i snitt 11 ord. Segmenter består derimot ikke kun av fullstendige setninger, men også av kulepunkter og overskrifter. Dermed vil gjennomsnittlig

Figur 5-3 Estimert antall segmenter i en 10-års periode



Kilde: Oslo Economics

Vi har estimert at omtrent 1,2 millioner segmenter vil kunne avleveres i rett filformat. Dette tallet er basert på antall segmenter de profesjonelle oversetterne har oppgitt at finnes i deres minner, i tillegg til segmentene som Amesto har i minnene til offentlige kunder. Ettersom det er stor usikkerhet knyttet til hvorvidt andre profesjonelle oversettere vil overlevere segmentene til sine offentlige kunder, har vi ikke inkludert et estimat på deres segmenter.

Videre antar vi at 20 prosent av segmentene ikke kan benyttes grunnet personvern hensyn eller liknende. Dermed anser vi 1,2 millioner som et minimumsestimat på antallet segmenter som kan innhentes i rett filformat.

Grunnet at parallelle tekster vil være vanskelige å innhente, er det knyttet betydelig usikkerhet til antallet segmenter som kan hentes inn gjennom disse. Vi vurderer det slik at det minimum vil være mulig å innhente 20 prosent av antallet segmenter som innhentes i rett filformat. Vi anser derfor 250.000 som et konservativt estimat på antallet segmenter.

BLD har allerede produsert 10.000 segmenter, som de ønsker at skal innlemmes i minnet til MT@EC. I satsingsnivå 3 legger vi til grunn at ytterligere segmenter innen forbrukerfeltet produseres, slik at et omfang av 100.000 segmenter (minimumsnivået EU-kommisjonen har estimert) produseres i satsingsnivå 3.

antall ord per segment være lavere enn gjennomsnittet per setning. Vi regner dermed med et snitt på 9 ord per segment.

Tabell 5-2 Forutsetninger i estimeringen

Satsingsnivå 0	Satsingsnivå 1	Satsingsnivå 2	Satsingsnivå 3
I dag finnes 733.081 segmenter	Vi har estimert at 300.000 segmenter kan samles inn årlig	Vi har estimert at 300.000 segmenter kan samles inn årlig	Vi har estimert at 300.000 segmenter kan samles inn årlig
	Vi har estimert at omtrent 1,2 millioner segmenter finnes i rett filformat i dag	Vi har estimert at omtrent 1,2 millioner segmenter finnes i rett filformat i dag	Vi har estimert at omtrent 1,2 millioner segmenter finnes i rett filformat i dag
		Vi antar at omtrent 250.000 segmenter finnes i andre filformat i dag	Vi antar at omtrent 250.000 segmenter finnes i andre filformat i dag
			Vi antar at 100.000 segmenter produseres
Totalt 0,7 mill. segmenter	Totalt 4,9 mill. segmenter	Totalt 5,2 mill. segmenter	Totalt 5,3 mill. segmenter

5.3 Vurdering av ikke-prissatte virkninger

I det følgende vurderer vi de ikke-prissatte virkningene. De ikke-prissatte virkningene er vurdert basert på pluss-minus metoden ihht. DFØs veileder for samfunnsøkonomiske analyser. Konsekvensen til en virkning, på en skala fra «----» (4 minuser) til «++++» (4 plusser), vurderes ut fra virkningens

betydning og omfang. *Betydningen av området som blir berørt for grupper i samfunnet eller for samfunnet som helhet vurderes enten som liten, middels eller stor. Omfanget vurderes som lite, middels eller stort ut fra i hvilken grad området blir berørt av tiltaket.*

Tabell 5-3 oppsummerer vurderingen vår av de ikke-prissatte virkningene.

Tabell 5-3: Vurdering av ikke-prissatte virkninger

Ikke-prissatte effekter	Betydning for samfunnet	Omfang			
		Satsingsnivå 1	Satsingsnivå 2	Satsingsnivå 3	Nynorsk-satsing
Kommunikasjon og informasjonsdeling på tvers av landegrensene	Stor	Lite positivt	Lite positivt	Middels positivt	Intet
Standardisering av språk	Middels	Middels positivt	Middels positivt	Middels positivt	Lite positivt
Språklig kvalitet	Middels	Lite negativt	Lite negativt	Lite negativt	Lite negativt
Misforståelser og støy i kommunikasjonen på tvers av landegrensene	Stor	Lite negativt	Lite negativt	Lite negativt	Intet/Lite positivt
Ivaretagelse av personvern og samfunnsikkerhet	Stor	Intet/Lite positivt	Intet/Lite positivt	Intet/Lite positivt	Intet/Lite positivt
Tilgjengeliggjøring av informasjon på nynorsk	Liten/Middels	Intet	Intet	Intet	Lite positivt
Konkurransen i markedet for maskinoversettelse	Middels	Intet/Lite negativt	Intet/Lite negativt	Intet/Lite negativt	Intet/Lite negativt

Mer kommunikasjon og informasjonsdeling på tvers av landegrensler

Mer kommunikasjon og informasjonsdeling på tvers av landegrensler vurderes å være av **stor betydning** for samfunnet.

Gjennom EU/EØS-samarbeidet kreves tett samhandling på tvers av land på mange områder. Norges evne til å utveksle informasjon og kommunisere med andre land er derfor av stor betydning for samfunnet. Mer effektiv informasjonsdeling vil også ha stor betydning for en rekke grupper i samfunnet. Eksempelvis kan velfungerende forbrukerportaler gjøre at rettighetene til privatpersoner styrkes på tvers av landegrensler.

Alle satsingsnivåene forventes å bidra til at flere tar i bruk maskinoversettelse og at antall oversettelser øker. Omfanget øker mest ved å gå fra satsingsnivå 0 til satsingsnivå 1, ved at segmenter fra en rekke nye fagområder inkluderes i corpus på norsk bokmål-engelsk. Vi forventer at både satsingsnivå 1 og 2 vil bidra til mer informasjonsdeling, men i et **lite positivt omfang**. Vi forventer ingen vesentlig forskjell på disse satsingsnivåene, men det kan bli noe mer informasjonsdeling i satsingsnivå 2 ettersom anvendelsen av verktøyet trolig blir noe bredere.

I satsingsnivå 3 forventer vi imidlertid at anvendelsesområdene kan utvides og at dette dermed bidrar til mer kommunikasjon og informasjonsdeling på nye områder. Begrunnelsen er at man med spesialproduserte corpora vil kunne styrke fagområder som eventuelt fortsatt har lavere oversettelses kvalitet. Vi forventer dermed at satsingsnivå 3 vil bidra til mer kommunikasjon og informasjonsdeling i et **middels positivt omfang**.

Når det gjelder nynorsksatsing forventer vi ikke at dette vil føre til mer informasjonsdeling utover at informasjon blir tilgjengelig på nynorsk.

Standardisering av språk

Standardisering av språk har vært et fokusområde for flere av de interne oversetterne vi har intervjuet. Konsistent språkbruk i administrative og juridiske tekster kan bidra til å unngå misforståelser og uklarheter. Oversettelsene som gjennomføres i offentlig sektor består i all hovedsak av administrative og juridiske tekster. Vi vurderer derfor at standardisering av språk er av **middels betydning** for samfunnet.

Vi forventer ingen vesentlig forskjell mellom satsingene når det gjelder standardisering av språk. Det kan imidlertid antas at et høyere satsingsnivå bidrar til mer bruk av verktøyet, også blant nye fagområder, og at omfanget på standardisering av språk dermed

øker. For alle de tre satsingsnivåene vurderer vi at omfanget på standardisering vil være **middels positivt**.

Når det gjelder nynorsksatsing legger vi til grunn at det vil påvirke standardisering av språk i et **lite positivt omfang**. Begrunnelsen er for det første at antall oversettelser fra EU-språk til nynorsk er begrenset. For det andre vil kvaliteten på oversettelsene til/fra nynorsk være relativt dårlig og bruken av verktøyet for oversettelser mellom nynorsk og engelsk forventes derfor å være begrenset (tilsvarende dagens oversettelse til bokmål).

Språklig kvalitet

Statistikkbasert oversettelse kan bidra til at språket blir mindre variert og rikt, samt at det kan føre til at kvaliteten på tekst blir dårligere ved at oversettelsen blir grammatisk feil, ved at feil ord benyttes eller ved at ord mangler i oversettelsen. Vi vurderer at språklig kvalitet er av **middels betydning** for samfunnet i denne sammenheng.

Alle satsingsforslagene vurderes å påvirke den språklige kvaliteten i et **lite negativt omfang**. Det kan imidlertid antas at et høyere satsingsnivå bidrar til mer bruk av verktøyet og dermed til en mer negativ innvirkning dersom verktøyet benyttes ukritisk. Vi mener imidlertid at det ikke kan forventes en stor nok forskjell til at det er verdt å skille på satsingsnivåene. En nynorsksatsing forventes også på samme måte å kunne bidra til at det blir lavere språklig kvalitet på oversettelsene til nynorsk.

Misforståelser og støy i kommunikasjonen på tvers av landegrensler

Dersom det oppstår misforståelser og støy i kommunikasjonen på tvers av land vurderer vi at det kan ha **stor betydning** for samfunnet.

Alle satsingsnivåer vil (sammenliknet med nullalternativet) kunne føre til mer støy og misforståelser, grunnet ukritisk bruk av verktøyet. Et høyere satsingsnivå vil innebære noe bedre språklig kvalitet, som kan bidra til mindre støy og misforståelser. Det kan imidlertid også tenkes at bedre språkkvalitet vil bidra til mer ukritisk bruk av verktøyet, og dermed bidra til mer støy og misforståelser.

Totaleffekten av høyere satsingsnivå er dermed usikker, men det forventes at forskjellene er av begrenset størrelse. Vi legger til grunn at alle satsingsnivåene vil påvirke graden av misforståelser i et **lite negativt omfang**. En nynorsksatsing vurderes å påvirke graden av misforståelser i liten grad ettersom utbredelsen av oversettelser mellom EU-språk og nynorsk antas å være liten også med en nynorsksatsing.

Ivaretagelse av personvern og samfunnssikkerhet

Personer har krav på at myndigheter ivaretar personvernet, og at de ikke deler personopplysninger med uvedkomne. Myndigheter har også ansvar for å tilse at sensitive opplysninger ikke blir offentliggjort. Ivaretagelse av personvern og samfunnssikkerhet vurderes å være av **stor betydning** for samfunnet.

Dersom uegnet bruk av gratis maskinoversettelse er utbredt kan det å gi virksomhetene tilgang på maskinoversettelse gjøre at man i større grad enn i dag ivaretar personvern og samfunnssikkerhet. Størrelsen på gevinsten avhenger av hvor stort problemet med bruk av Google Translate og Bing Translate er opprinnelig.

Dersom bruken av uegnede oversettelsesverktøy er et lite utbredt problem i offentlige virksomheter, vil en satsing på MT@EC heller ikke påvirke ivaretagelsen av personvern og samfunnssikkerhet i særlig grad. Ettersom omfanget av bruken er usikker, vurderer vi at satsingsnivåene påvirker ivaretagelsen av personvern og samfunnssikkerhet i et omfang som kan variere fra **intet** til **lite positivt**.

Tilgjengeliggjøring av informasjon på nynorsk

Ved å etablere nynorsk i MT@EC vil det bidra til at flere dokumenter og tekster oversettes til nynorsk. Dette har en positiv effekt ved at mer informasjon blir tilgjengelig på nynorsk. Vi vurderer at tilgjengeliggjøring av informasjon på nynorsk er av **liten** til **middels betydning** for samfunnet.

Satsingsforslag 1 til 3 vil ikke bidra til å gjøre mer informasjon tilgjengelig på nynorsk. Når det gjelder en nynorsksatsing i MT@EC forventer vi at det vil påvirke tilgjengeliggjøring av dokumenter på nynorsk i et **lite positivt omfang**. Begrunnelsen er for det første at det allerede finnes maskinoversettelsesverktøy, eksempelvis Nyno, som kan benyttes til oversettelser mellom norsk og nynorsk. For det andre er omfanget av corpora på nynorsk såpass begrenset at kvaliteten på oversettelsene til nynorsk kan forventes å være relativt dårlig.

Konkurranse i markedet for maskinoversettelse

Konkurranse i markedet for maskinoversettelse er viktig for at det skal utvikles nye og bedre løsninger innen maskinoversettelse. Derimot er markedet for maskinoversettelse relativt lite i et samfunnsperspektiv. Vi vurderer derfor at det å unngå konkurranse- vridende effekter er av **middels betydning** for samfunnet. Dersom språkressursene MT@EC benytter tilgjengeliggjøres for forsknings- og utviklermiljøer gjennom eksempelvis Språkbanken, vurderer vi at MT@EC ikke vil skape betydelige negative effekter på konkurransen i markedet for maskinoversettelse. Videre vurderer vi at de konkurransemessige

effektene ikke vil være veldig forskjellige i ulike satsingsnivåene.

Vi vurderer derfor at effekten på konkurransen i markedet for maskinoversettelse er **intet** til **lite negativt**.

5.3.1 Oppsummert vurdering av de ikke-prissatte virkningene

Vurderingen av de ikke-prissatte virkningenes betydning og omfang bestemmer til sammen virkningens konsekvens. I Tabell 5-4 presenterer vi virkningenes konsekvens på en nidelt skala fra ++++ (meget stor positiv konsekvens) til - - - - (meget stor negativ konsekvens).

Tabell 5-4: Virkningenes konsekvens

Virkning	1	2	3	Nynorsk
Kom. og informasjonsdeling på tvers av landegrenser	+	++	+++	0
Standardisering av språk	++	++	++	+
Språklig kvalitet	0/-	0/-	0/-	0/-
Misforståelser og støy i kommunikasjonen på tvers av landegrenser	-/--	-/--	-/--	-
Ivaretagelse av personvern og samfunnssikkerhet	0/+	0/+	0/+	0/+
Tilgjengeliggjøring av informasjon på nynorsk	0	0	0	0/+
Konkurranse i markedet for maskinoversettelse	0/-	0/-	0/-	0/-

Samlet sett vurderer vi at det er små forskjeller på virkningene man oppnår med ulike satsingsnivå. De viktigste nytteeffektene av en satsing på MT@EC er mer kommunikasjon og informasjonsdeling på tvers av landegrenser. Her vurderer vi at et økt satsingsnivå vil bidra til at man får større gevinster ettersom det forventes at bruken av verktøyet da vil øke og ved at man utvider antall anvendelsesområder for verktøyet.

En satsing på nynorsk vurderes å gi små effekter ettersom omfanget av oversettelser mellom nynorsk og EU-språk er såpass begrenset. Standardisering av språk og mer konsistent språkbruk er en annen vesentlig effekt man oppnår ved å satse på MT@EC. Her vurderer vi imidlertid at forskjellen i hva man oppnår ikke vil være stor mellom de ulike satsingsnivåene.

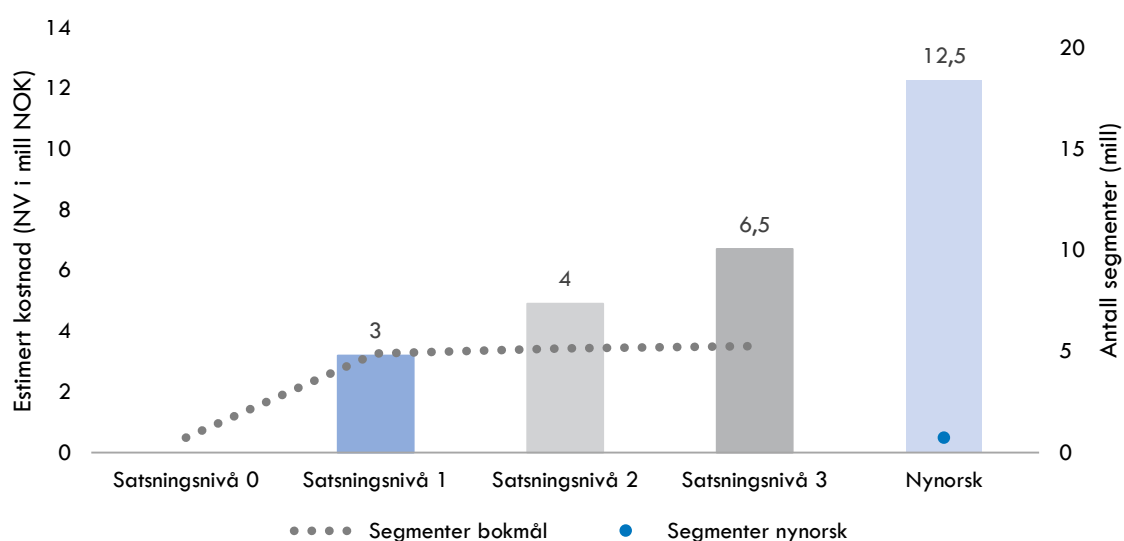
5.4 Vurdering av prissatte virkninger ved en satsing på MT@EC

Figur 5-4 gir en oversikt over estimerte kostnader ved de ulike satsingsnivåene. Antall segmenter er oppgitt som en stiplet linje i samme diagram. Figuren illustrerer at kostnader fortsetter å stige ved økt satsingsnivå, til tross for at antallet segmenter stagnerer. Årsaken til dette er at kostnaden per segment øker betraktelig ved ytterligere satsing. Denne resultatet fremstår som

rimelig. I satsingsnivå 2 må eksempelvis innhentede filer konverteres, noe som ikke er nødvendig i satsingsnivå 1. I satsingsnivå øker kostnaden per segment ytterligere, ettersom tilleggssegmentene må produseres fra bunnen av.

Grunnet at virksomhetene i liten grad har nynorsk-segmenter, er kostnaden for en nynorsksatsing større enn for bokmålsatsingen. For å produsere et nynorsk-corpus på samme størrelse som dagens bokmål-corpus (nullalternativet), estimerer vi kostnadene til å være 12,5 millioner kroner i netto nåverdi.

Figur 5-4 Estimert kostnad (NNV i mill. kroner)



Kilde: Oslo Economics

I kostnadsberegningen legger vi til grunn at prisen til et årsverk lønn pluss sosiale utgifter. Vi benytter gjennomsnittslønn for statlige ansatte i 2014 (SSB-tabell 08702), som er 526.300 kroner per årsverk. Inkludert sosiale utgifter på 30 prosent blir kostnaden per årsverk 684.190 kroner. I netto nåverdi-beregningen antar vi videre en lønnsvekst på 1,3

Tabell 5-5.

Interessentene anslår det som lite tidkrevende å legge til rette for kontinuerlig fremtidig avlevering av segmenter. Dersom de profesjonelle oversetterne fortløpende kopierer minnefiler som kan overleveres antar vi dermed at denne kopieringen utgjør et ukesverk per virksomhet med interne oversettere. Dette er estimater basert på uttalelser fra eksempelvis Justisdepartementet og UD's oversettelsesavdeling og at visse dokumenttyper er enkle å «sile ut» som uproblematiske, ettersom de ikke inneholder sensitiv informasjon. Dersom slik «utsiling» er mulig, vil det ikke

prosent¹⁵ og en diskonteringsfaktor på 4 prosent i tråd med DFØ's veileder.

Basert på uttalelser fra dybdeintervjuene har vi anslått antall årsverk til de ulike trinnene i satsingen på MT@EC. En samlet oversikt over antakelsene er gitt i

kreve mye ressurser å avlevere slike uproblematiske dokumenter og tilhørende segmenter.

I tillegg til utgifter i virksomhetene antar vi årlig ressursbruk sentralt for innsamling av corpora på 0,25 årsverk. I tillegg antar vi at etableringen av en slik avleveringsordning ikke vil overstige et årsverk. Estimert kostnad per segment ved kontinuerlig avlevering blir gitt disse forutsetningene omtrent 0,6 kroner per segment.

Sammenliknet med kontinuerlig avlevering av segmenter som produseres i fremtiden, anser

¹⁵ Ifølge DFØ's veileder skal lønnsveksten fra nyeste perspektivmelding benyttes. Dermed benyttes en lønnsvekst på 1,3 prosent, i tråd med Perspektivmeldingen fra 2013.

interessentene våre det som mer krevende å foreta en gjennomgang av eksisterende minner.

Eksempelvis estimerer Statens vegvesen at omtrent 5-10 prosent av deres oversettelsesminne kan inneholde personopplysninger eller forretningshemmeligheter. Det kreves en gjennomgang av alle segment for å fjerne problematiske opplysninger. Andre virksomheter, eksempelvis UD's oversettelsesavdeling, har organisert segmentfilene i underkategorier, der noen kategorier er mer uproblematiske enn andre.

Det finnes få virksomheter med interne profesjonelle oversettere, og dermed blir ressursbehovet for gjennomgang av segmentfiler begrenset. Vi anslår dermed en ressursbruk på 1,5 årsverk, som tilsvarer 3-4 månedsverk i snitt per virksomhet med interne oversettere. Estimert kostnad per segment ved innsamling av eksisterende segmenter blir dermed omtrent 0,9 kroner per segment.

Informasjon innhentet gjennom dybdeintervjuene tyder på at det vil være svært tidkrevende å finne parallelle tekster i andre filformat. Årsaken til dette vil være at slike tekster i et samlet på ett sted og at det kan være utfordrende å finne både originaldokumentet i tillegg til oversettelsen. SSB

Tabell 5-5 Estimering av kostnader

	Satsingsnivå 1	Satsingsnivå 2	Satsingsnivå 3
Årlig avlevering:	1 årsverk til etablering av ordningen, 0,25 årsverk i drift (1 ukesverk per virksomhet med interne oversettere og 1 månedsverk sentralt)	1 årsverk til etablering av ordningen, 0,25 årsverk i drift (1 ukesverk per virksomhet med interne oversettere og 1 månedsverk sentralt)	1 årsverk til etablering av ordningen, 0,25 årsverk i drift (1 ukesverk per virksomhet med interne oversettere og 1 månedsverk sentralt)
Avlevering eksisterende segmenter:	1,5 årsverk til gjennomgang, screening og avlevering av segmenter	1,5 årsverk til gjennomgang, screening og avlevering av segmenter	1,5 årsverk til gjennomgang, screening og avlevering av segmenter
Avlevering og konvertering andre filformat:		2,5 årsverk til gjennomgang og avlevering av parallelle dokumenter og konvertering av filformat	2,5 årsverk til gjennomgang og avlevering av parallelle dokumenter og konvertering av filformat
Produksjon av corpora:			100 000 segmenter, 9 ord per segment i snitt og 2 kr per ord

karakteriserer eksempelvis en innsamling av parallelle Word-filer som umulig eller veldig ressurskrevende.

Videre må tid til konvertering påberegnes. Vi anslår derfor en ressursbruk på omtrent 2,5 årsverk for 250.000 segmenter, noe som fører til en kostnad på omtrent 7 kroner per segment.

For produksjonen av corpora legger vi til grunn en gjennomsnittlig markedspris på 2 kr per ord og 9 ord per segment. Dette fører til en pris på 18 kroner per segment og en samlet kostnad på 1,8 millioner kroner for å produsere 100.000 segmenter.

Ettersom virksomhetene rapporterer om svært få oversettelser til/fra nynorsk, legger vi til grunn at det heller ikke finnes segmenter hos virksomhetene som kan samles inn. Derfor oppstår ingen kostnader knyttet til innsamling av nynorsk, verken i form av kontinuerlig avlevering eller i form av innhenting av eksisterende segmentfiler og parallelle tekster. For å nå dagens antall bokmål-segmenter kreves dermed at i underkant av 700.000 segmenter produseres. Ved en kostnad på 18 kroner per segment koster denne nynorsk-satsingen dermed omtrent 12,3 millioner kroner.

5.5 Usikkerhet

Våre analyser av virkningene bygger på ulike antakelser som det er knyttet usikkerhet til. I dette avsnittet vil vi gjøre rede for viktige usikkerhetsmomenter og hvordan disse påvirker virkningene til en satsing på MT@EC.

5.5.1 Omfang av tekst som kan benyttes i corpus

Dette er usikkerhet knyttet til antakelsene våre om hvor mye tekst som kan benyttes i corpus.

En stor andel av segmentene inneholder sensitive opplysninger, og kan dermed ikke overleveres. Dessuten er eierskap til segmentfilene et juridisk uavklart spørsmål. Dersom kjøpere av oversettelsestjenester er juridiske eiere også til segmentminnene, vil antallet segmenter som kan innhentes øke betraktelig. Videre er det knyttet betydelig usikkerhet til omfanget av segmenter som kan innhentes gjennom andre filformat.

I beregningene våre har vi imidlertid lagt til grunn relativt konservative anslag på omfanget av tekst som innhentes/produseres i de ulike satsingsnivåene. Vi vurderer derfor ikke at dette usikkerhetsmomentet har vesentlig betydning for vår analyse.

5.5.2 Kvaliteten på maskinoversettelsene

Kvaliteten på maskinoversettelsene vil også påvirkes av kvaliteten på oversatte corpora som avleveres. Dersom den språklige kvaliteten på corpuset er lav, vil også maskinoversettelsene bli dårligere.

Kvaliteten på maskinoversettelsene er en viktig faktor for hvorvidt virksomhetene vurderer at maskinoversettelse vil gi gevinster.

Det er grunn til å tro at kvaliteten på tekstene som samles inn fra de profesjonelle oversetterne i satsingsnivå 1 er relativt god. Det er imidlertid større usikkerhet knyttet til hvor god kvaliteten på tekster som samles inn fra øvrige parter vil være. Dersom man fokuserer på å samle inn offisielle dokumenter tror vi imidlertid at det vil bidra til relativt god språklig kvalitet på corpora.

5.5.3 Omfanget på bruken av verktøyet

Det er også usikkerhet knyttet til i hvilken grad verktøyet blir tatt i bruk. Dette vil være helt avgjørende for gevinstrealiseringen.

Blant **profesjonelle oversettere** vil det grunnleggende synet på maskinoversettelse i stor grad kunne påvirke hvorvidt de ønsker å ta i bruk verktøyet. Dersom en oversetter eksempelvis er prinsipielt imot maskinoversettelse er det også lite trolig at oversetteren vil benytte seg av MT@EC.

Når det gjelder **webtjenesten til MT@EC** vil kvaliteten på maskinoversettelsen være avgjørende for hvorvidt offentlige ansatte vil ta i bruk tjenesten. For mange er det også vesentlig at oversettelsene benytter terminologien som anvendes på deres fagområde for at de skal bruke en slik oversettelsestjeneste. Undersøkelsene våre tyder på at oversettelsesverktøy som for eksempel Google Translate er relativt utbredt. Dersom man får en tilsvarende tjeneste av høyere kvalitet med mer tilpasset terminologi kan det dermed tyde på at det vil være en viss etterspørsel etter en slik tjeneste. I hvilken grad man klarer å oppnå en tilfredsstillende kvalitet for de statlige ansatte i de ulike satsingsforslagene er imidlertid noe usikkert.

Vi vurderer at det er knyttet minst usikkerhet til **bruken av verktøyet i nettportaler**, ettersom integrering av verktøyet foretas fra sentralt hold i EU-kommisjonen, i samarbeid med nasjonale myndigheter. ODR-plattformen er et eksempel på dette. Med videreføringen av MT@EC i CEF.AT har EU som ambisjon å gjøre samtlige digitale tjenester flerspråklige. Det foreligger konkrete planer om integrering av MT@EC i minst seks nye webportaler i tillegg til de fem webportalene hvor MT@EC allerede er integrert. Ved integrering av MT@EC i disse webportalene vil MT@EC anvendes til oversettelse av de som benytter disse webportalene. Dermed er det mindre usikkerhet knyttet til bruken av MT@EC i disse verktøyene.

5.6 Samlet vurdering og anbefaling

Den mest vesentlige nyttevirkningen av en satsing på MT@EC er mer kommunikasjon og informasjonsdeling på tvers av landegrensene. Vi forventer at et økt satsingsnivå vil bidra til større gevinster på dette området ettersom det forventes at den generelle bruken av verktøyet øker og at man kan anvende MT@EC på stadig flere områder.

Standardisering av språk og mer konsistent språkbruk er andre vesentlige virkninger man kan oppnå ved å satse på MT@EC. Her vurderer vi imidlertid at forskjellen i hva man oppnår ikke vil være stor mellom de ulike satsingsnivåene. For de øvrige ikke-prissatte virkningene vurderer vi at det er små forskjeller i hva man oppnår med et høyere satsingsnivå.

Vi estimerer at kostnaden for en minimumssatsing vil være 3 millioner kroner i netto nåverdi over en 10-årsperiode. Vi vurderer at de positive effektene MT@EC medfører i form av økt informasjonsutveksling på tvers av landegrensene og høyere kvalitet på oversettelser kan forsvare en satsing utover dagens nivå.

Som illustrert i Figur 5-4, vurderer vi at en satsing utover satsingsnivå 1 vil gi en betydelig relativ kostnadsøkning, samtidig som økningen i omfanget av corpora er liten. Vi vurderer på bakgrunn av dette at satsingsnivå 1 fremstår som mest hensiktsmessig. Et vesentlig poeng her er at satsingsnivå 1 ikke utelukker muligheten for en videre satsing i fremtiden. Dermed

kan satsingsnivå 2 og 3 anses som en opsjoner. Med satsingsnivå 1 vil man trolig få en kvalitetsforbedring i oversettelsene på en rekke fagområder. Etter en vurdering av gevinstrealiseringen som følge av satsingsnivå 1, vil det være mulig å investere ytterligere i corpora-innsamling og produksjon av corpora ved behov. Ved produksjon av corpora kan man da prioritere de områdene hvor man behovet for kvalitetsforbedringer er stort, eksempelvis på forbrukerområdet.

En satsing på nynorsk kan foretas uavhengig av avgjørelsen om bokmålsatsingen. Ettersom en nynorsksatsing krever at et stort volum segmenter produseres, vil en slik satsing bli kostbar. Sammenliknet med det anbefalte alternativet i bokmålsatsingen, satsingsnivå 1, er den estimerte kostnaden til nynorsksatsingen nesten fire ganger så stor. Basert på svarene til virksomhetene og de ansatte i spørreundersøkelsene tillegges nynorskoversettelser liten verdi. En høy kostnad, kombinert med lavt rapportert behov for nynorskoversettelser gjør at det fremstår som lite hensiktsmessig å investere i produksjonen av et nynorsk-corpus. En slik satsing vil likevel kunne være forsvarlig dersom muligheten til nynorskoversettelser tillegges svært høy egenverdi. Samlet sett vurderer vi at det ikke er grunnlag for å anbefale at et nynorsk-corpus produseres basert på innspillene vi har fått fra virksomhetene.

6. Vedlegg

Tabell 6-1: Behovet for oversettelse til norsk i EU portaler/nettsteder

Virksomhet	EU-portaler/-nettsteder som benyttes	Behov for oversettelse til norsk bokmål
Arbeids- og sosialdepartementet	Hjemmesidene til EU-kommisjonen: ec.europa.eu	I noen grad
Barne- og likestillingsdepartementet	Benytter blant annet CircaBC, IMI – Internal Markets Information System, ODR	Ja – Arbeidet med oversettelse til norsk bokmål i ODR, vurderes også å være nyttig på enkelte andre områder
Brønnøysundregistrene	IMI, hjemmesidene til EU-kommisjonen: ec.europa.eu med mange undersider, samt INEA-programmer, ISA2-Programmer, Horizon2020, etc.	Ja
Datatilsynet	Benytter CircaBC og andre samarbeidsportaler.	Ja - Nyttig med oversettelse til/fra norsk bokmål i samarbeidsportalene som benyttes.
Direktoratet for økonomistyring	Ingen	Nei
Direktoratet for byggkvalitet	Benytter Rapex, ICSMS og CircaBC.	I noen grad. For tekster som brukes innhentes utenfra for intern bruk vil det være mindre behov for oversettelse, behovet for rask og enkel oversettelse vil være større ved innhenting av tekster som skal offentliggjøres.
Direktoratet for samfunnssikkerhet og beredskap	Benytter CECIS, ERCC og GDACS.	Nei - Vurderer ikke at det er behov for oversettelse til norsk bokmål.
Ecolabel	EU Ecolabel og øvrige EU-nettsteder	Nei – Kan lese dokumenter på engelsk eller dansk
Finansdepartementet	Benytter flere av EUs nettportaler og nettsteder	Nei
Finanstilsynet	Benytter portalene til EUs finanstilsyn (EBA, ESMA og EIOPA) og EUR-Lex.	<i>Ikke mottatt svar</i>
Forbrukerombudet	Benytter Consumer Protection Cooperation System (CPCS) til behandling av europeiske forbrukersaker	Ja – Forbrukerombudet mener at det hadde vært til stor hjelp for oss dersom dette ble implementert i CPCS. Her ligger det allerede inne oversettelse mellom alle de offisielle EU-språkene.

Havarikommisjonen	Benytter felles europeiske databasesystemer innenfor de ulike transportområdene: European Aviation Safety Agency, European Railway Agency, European Marine Casualty Information Platform m.fl.	Usikker - Det er vanskelig å si om verktøyet trenger å implementeres direkte i portalene, eller om det er tilstrekkelig at vi har tilgang til å få tekst oversatt utenom.
Helse- og omsorgsdepartementet	Hjemmesidene til EU-kommisjonen: ec.europa.eu	Nei
Integrerings- og mangfoldsdirektoratet	Benyttes ikke	Nei
Innovasjon Norge	Rapportering for Norway Grants, ellers lite.	I noen grad – Potensielt oversettelse behov for Norway Grants
Justisdepartementet	TED (Tenders Electronic Daily), eJustice	I noen grad – Doffin oversetter allerede dokumenter til engelsk som publiseres i TED
Kripes	<i>Har ingen oversikt</i>	I noen grad – Integrering i politiets egne saksbehandlingsverktøy for internasjonale saker og etterretningsinformasjon kunne vært nyttig, men krever at systemet ivaretar politiets krav til informasjonssikkerhet.
Kunnskapsdepartementet	EU-kommisjonens systemer som benyttes av DG EAC og Research, samt av noen byråer innenfor fagområdene. I mindre grad Europaparlamentet og Det europeiske råds systemer.	Nei
Mattilsynet	EUR-Lex, CIRCABC, m.fl.	Ja
Miljødirektoratet	Flere av EU-kommisjonens nettsider spesielt http://www.echa.europa.eu/ og EU-kommisjonens høringside	I noen grad – kunne vært nyttig og oversette veiledere, høringer som legges ut annet materiale.
Nasjonale kommunikasjonsmyndighet (NKOM)	Benytter i noen grad EU-portaler som EUR-Lex, EU-domstolens portal, EU-kommisjonens portal og EFTAS portal.	I noen grad – Informasjonen finnes allerede på språkene innenfor EU, men litt gevinst dersom informasjonen også blir tilgjengeliggjort på norsk bokmål.
Nasjonalt organ for kvalitet i utdanningen (nokut)	EU-portaler for yrkeskvalifikasjonsrammeverket og andre saksbehandlingsverktøy innen feltet godkjenning utenlandsk utdanning.	Usikker - vanskelig å si siden om det er behov uten å vite hvordan et slikt program vil være bygd opp og hvor mye gjennomgang av den maskinoversatte teksten som kreves.

Nærings- og fiskeridepartementet	NANDO (nettsted for notifikasjon av tekniske kontrollorganer), TRIS (database for melding av tekniske krav til varer). IMI inneholder flere moduler, og benyttes for SOLVIT (samarbeid mellom EØS-statene for å løse problemer borgere og bedrifter møter som følge av uriktig anvendelse av EU/EØS-retten) og notifikasjon av nasjonale krav for utøvelse av tjenestevirksomhet/etablering.	I noen grad – Det er avhengig av hvor gode oversettelser systemet gir. Notifikasjon av nasjonalt regelverk for varer (TRIS) og tjenester (IMI) innebærer oversettelse av norske lover/forskrifter til engelsk. Denne oversettelsen foretas ikke av NFD, men av ansvarlig departement/etat. Maskinoversettelse vil kunne gjøre dette arbeidet mindre ressurskrevende.
Olje- og energidepartementet	Ingen portaler benyttes	Nei
Petroleumstilsynet	Benyttes i liten grad	Nei
Samferdselsdepartementet	Benytter hjemmesidene til EU-kommisjonen: ec.europa.eu	Nei
Skattedirektoratet	Benytter hjemmesidene til EU-kommisjonen: ec.europa.eu og ulike nasjoners sider innenfor skatteområdet.	Nei
Riksrevisjonen	Benytter TED	Usikker – vanskelig å uttale seg uten å kjenne verktøyet nærmere
Utlendingsdirektoratet	Benytter i stor grad ulike EU-portaler/nettsteder. Eks: EUR-LEX, CIRCABC, CURIA og nettsiden til EU-kommisjonen.	Ja – Det kan være et nyttig verktøy ved bruk av EU-portalene/nettstedene som benyttes av UDI

oslo**economics**

www.osloeconomics.no

post@osloeconomics.no
Tel: +47 21 99 28 00
Fax: +47 96 63 00 90

Besøksadresse:
Dronning Mauds Gate 10
0250 Oslo

Postadresse:
Postboks 1540 Vika
0117 Oslo