

Kartlegging av behovet for automatisk oversettelse i statlig sektor

*Workshop om automatisk oversettelse - European
Language Resource Coordination (ELRC)*

8. juni 2016

oslo**economics**

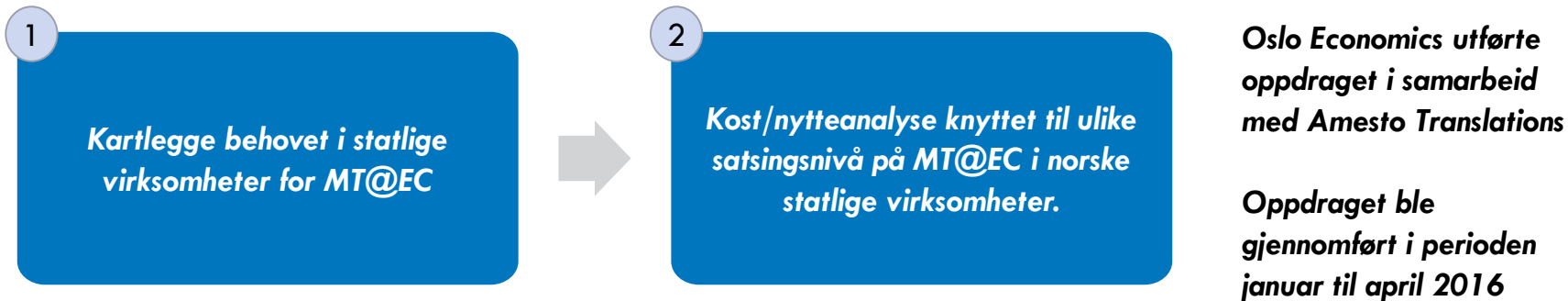


Kartlegging av behovet for automatisk oversettelse i statlig sektor

1. **Innledning**
2. MT@ECs målgrupper og
bruksområder
3. Vurdering av behovet for MT@EC
4. Kost/nytteanalyse fremtidig satsing



Vårt oppdrag for KMD



Informasjonskilder

Innhentet informasjon:

Dybdeintervjuer

- Dybdeintervjuer med DG Translate, DG Connect og den svenske oversetteravdelingen i EU-kommisjonen
 - Spyridon Pilos og Markus Foti i DG Translate
 - Kimmo Rossi i DG Connect
 - Tomas Backlund, svenske oversetteravdelingen i EU
- 12 intervjuer med ulike departement og etater – hovedvekt på virksomheter med interne oversettere
- Intervju med forsker Trond Trosterud ved Universitetet i Tromsø.

Spørreundersøkelse

- Spørreundersøkelse hos **etater og departementer** for å utvide utvalget.
 - Undersøkelsen ble sendt ut til 36 virksomheter, og vi mottok svar fra 30.
- Spørreundersøkelse blant **ansatte** i Mattilsynet, BLD, Innovasjon N., Forbruker Europa, Vegvesenet, Sjø.dir. og Finanstilsynet og UDI – 45 svar (svarprosent på 44 %)

Utredningen har også bygget på skriftlig informasjon om MT@EC/CEF.AT primært fra EU-kommisjonen

Kartlegging av behovet for automatisk oversettelse i statlig sektor

1. Innledning
2. MT@ECs målgrupper og bruksområder
3. Vurdering av behovet for MT@EC
4. Kost/nytteanalyse fremtidig satsing



Dagens anvendelser av MT@EC og målgrupper i Norge

Anvendelser

Webtjeneste for oversettelse av dokumenter

Målgrupper i Norge

- Primært offentlig ansatte og saksbehandlere.
- Profesjonelle oversettere

Anvendelse i EUs online tjenester/plattformer

- Forbrukere, arbeidstakere og arbeidsgivere, etc.

Dagens og fremtidige anvendelser i EUs online tjenester/plattformer

- MT@EC er i dag implementert i en rekke EU-webportaler (IMI, Solvit, Eur-Lex, CircaBC og TED), men foreløpig oversettes det ikke til norsk i disse webportalene.
- Konkrete planer om å integrere MT@EC i flere webportaler på sikt som en del av CEF.AT (Europeana, Open Data Portal, EESSI, ODR og e-justice).

Kartlegging av behovet for automatisk oversettelse i statlig sektor

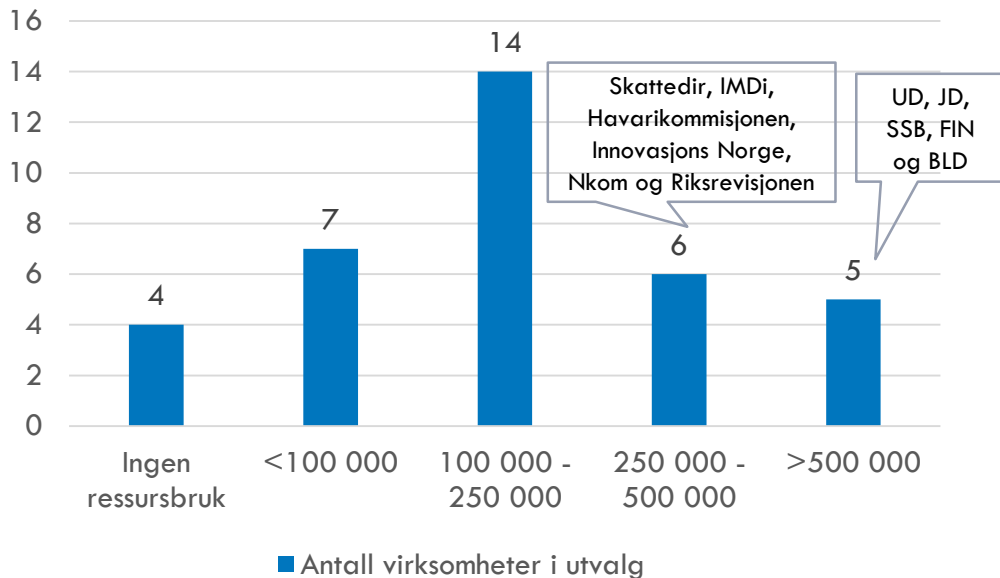
1. Innledning
2. MT@ECs målgrupper og
bruksområder
3. **Vurdering av behovet for MT@EC**
4. Kost/nytteanalyse fremtidig satsing



Behovet for oversettelse i statlige virksomheter er varierende

Kostnader til innkjøp av oversettelsestjenester

Ressursbruk i kr eks mva, antall virksomheter per kategori



- 85 prosent av virksomhetene har en ressursbruk under 250 000 kr årlig til eksterne oversettere.
- Noen virksomheter skiller seg ut når det gjelder oversettelse:
 - UD har klart størst ressursbruk knyttet til innkjøp av oversettelse med ca. 4,8 mill. kr årlig
 - Justisdepartementet har også betydelige utgifter (1,2 mill. kr årlig)

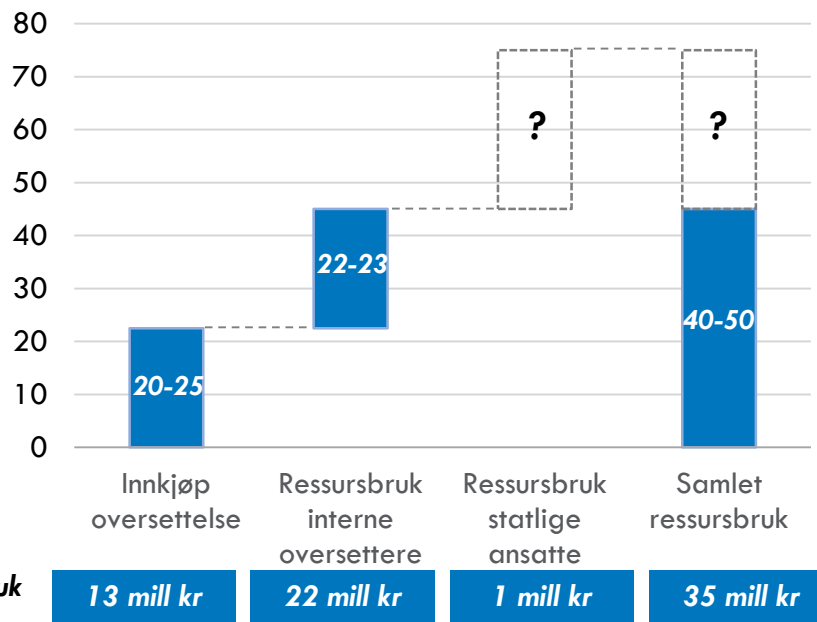
Gjennomsnittlig ressursbruk er rundt 370 000 kr årlig per virksomhet i utvalget

Flere virksomheter i staten har interne oversettere

Virksomhet	Antall årsverk	Type oversettelser
UD - EØS-oversettelse	19	<ul style="list-style-type: none">• Oversetter rettsakter i form av de delene av EUs regelverk som innlemmes i EØS-avtalen.
UD - Seksjon for oversettelse	6,5	<ul style="list-style-type: none">• Oversetter eksempelvis folkerettslige avtaler, politiske dokumenter og tekster, strategier og handlingsplaner for UD.
Statens vegvesen	2,5	<ul style="list-style-type: none">• Oversetter eksempelvis regler og forskrifter, veg- og informasjonsskilt, powerpoint-presentasjoner, brev og elektroniske løsninger/søknadsskjema.
Sjøfartsdirektoratet	2,1	<ul style="list-style-type: none">• Oversetter med forskriftsendringer og internasjonalt regelverk.
Finanstilsynet	1	<ul style="list-style-type: none">• Oversetter årsmeldingen, rapporten «Finansielt utsyn», lover, forskrifter og rundskriv til engelsk.
Stortinget	0,5	<ul style="list-style-type: none">• Jobber primært med kvalitetskontroll og språkvask av taler og tekster til hjemmesider.
Totalt	31,6	

Dagens ressursbruk til oversettelse er en indikator på behovet for å ta i bruk maskinoversettelse

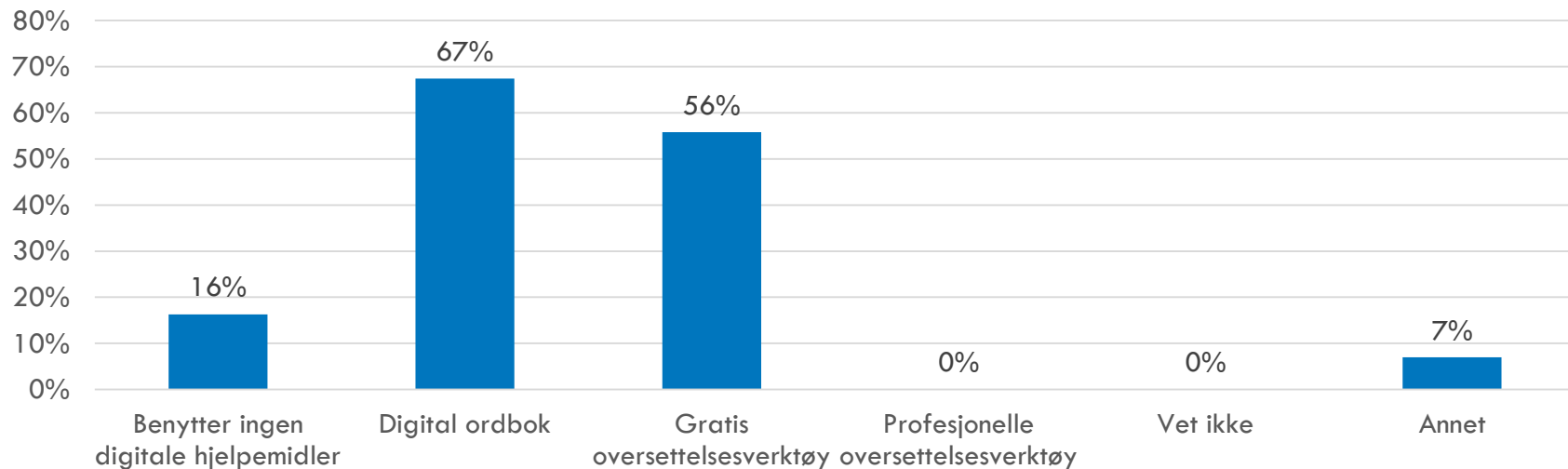
Estimert årlig ressursbruk til oversettelse i mill. kr eks mva



- Vi estimerer at samlet ressursbruk i statlig sektor til oversettelse er 40-50 millioner kroner årlig
- Estimater er basert på Amestos markedsandel og inntekter fra ekstern oversettelsesvirksomhet, samt undersøkelse av ressursbruk blant interne oversettere i statlige virksomheter.
- Også trolig betydelig ressursbruk knyttet til oversettelser som gjennomføres av ansatte i statlige virksomheter, men vi gir ingen konkrete anslag på denne størrelsen som følge av usikkerhet.

Bruken av digitale hjelpemidler til oversettelse er relativt utbredt blant statlige ansatte

Hvilke digitale hjelpemidler benytter du i dag når du oversetter tekster?
– andel av utvalget, 43 svar



Vurdering av behov for MT@EC blant interne oversettere

	Oversettelsesbehov
Språk	<ul style="list-style-type: none">• Oversetter i all hovedsak mellom norsk og engelsk
Type tekster	<ul style="list-style-type: none">• Juridiske tekster: Lover, forskrifter og rundskriv• Andre tekster: Hjemmesider, taler, etc.
Fagspråk	<ul style="list-style-type: none">• Mye juridisk og administrativt språk, men oversetter også hjemmesider, taler, etc.
Bruk av digitale hjelpemidler	<ul style="list-style-type: none">• De fleste interne oversetterne i staten benytter et såkalt «CAT»-verktøy

Vurdering av behov for MT@EC

- MT@EC vurderes å gi begrensede effektivitetsgevinster utover det de allerede har realisert ved å ta i bruk CAT-verktøy.
- Gevinsten avhenger i stor grad av kvaliteten på oversettelsen og hvor stort minne virksomheten har fra før.
- Gevinst forutsetter effektiv integrering i dagens CAT-verktøy.

Samlet sett vurderes behovet interne oversettere har for MT@EC som relativt begrenset.

Vurdering av behov for MT@EC i statlige virksomheter

Øversettelsesbehov

Språk

- Primært behov for å oversette mellom norsk og engelsk, men enkelte har også behov for å oversette til andre EU-språk .

Type tekster

- Stor bredde i type tekster som oversettes.

Fagspråk

- Behov for fagspesifikt språk og at maskinoversettelsen benytter de relevante begrepene.

Bruk av digitale hjelpemidler

- Google translate og digital ordbok anvendes av mange.

Vurdering av behov for MT@EC

- Litt over halvparten av virksomhetene uttaler at MT@EC vil gi virksomheten gevinster.
- Utbredt bruk av gratis oversettelsesverktøy som Google Translate kan også antyde at det er et behov for MT@EC.
- Virksomhetene er opptatt av at oversettelsen må ha relativt høy kvalitet og anvende relevant fagterminologi for at det skal gi de gevinster.
- Primært gevinst i form av at mer oversettes – forventes ingen kostnadsreduksjon som følge av MT@EC.

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 oversettelsen er tilpasset det aktuelle fagområdet.

Vurdering av behov for MT@EC i statlige virksomheter

Utvalgte uttalelser om behovet for MT@EC

- «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 IMI, epost firmapostmottak, informasjonsbehandling og saksbehandling.» - Brønnøysundregistrene
- «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.» - NAV
- «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.» - UDI
- «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.» - Finanstilsynet

Kartlegging av behovet for automatisk oversettelse i statlig sektor

1. Innledning
2. MT@ECs målgrupper og
bruksområder
3. Vurdering av behovet for MT@EC
4. **Kost/nytteanalyse fremtidig satsing**

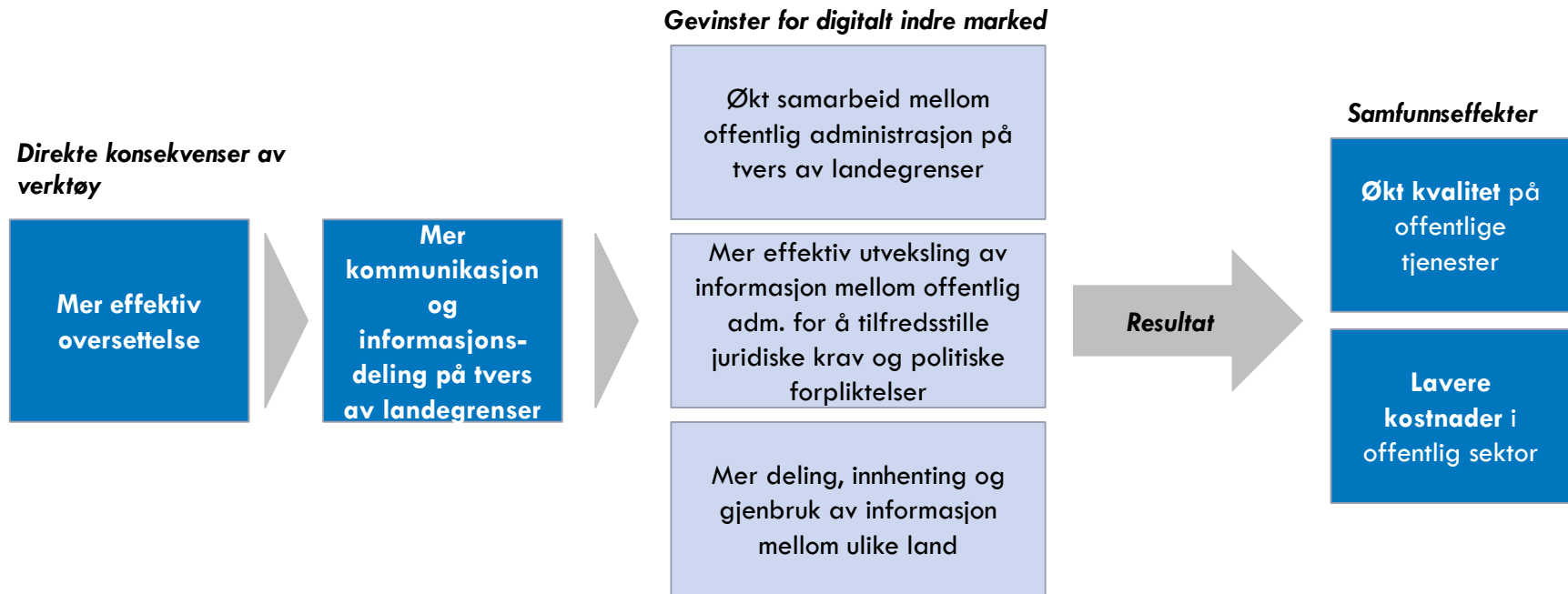


Kost/nytteanalyse fremtidig satsing på MT@EC

Satsningsnivå 0	Satsningsnivå 1	Satsningsnivå 2	Satsningsnivå 3	Nynorsksatsing
<ul style="list-style-type: none"> Ingen endring i antall segmenter 	<ul style="list-style-type: none"> Kontinuerlig avlevering av segmenter 	<ul style="list-style-type: none"> Kontinuerlig avlevering av segmenter 	<ul style="list-style-type: none"> Kontinuerlig avlevering av segmenter 	
	<ul style="list-style-type: none"> Innhenting av eksisterende filer i rett filformat 	<ul style="list-style-type: none"> Innhenting av eksisterende filer i rett filformat 	<ul style="list-style-type: none"> Innhenting av eksisterende filer i rett filformat 	
		<ul style="list-style-type: none"> Innhenting av parallelle tekster i andre filformat 	<ul style="list-style-type: none"> Innhenting av parallelle tekster i andre filformat 	
			<ul style="list-style-type: none"> Produksjon av corpora 	<ul style="list-style-type: none"> Produksjon av corpora

Nytte- og kostnadsvirkninger av en satsing på MT@EC

Viktigste nytteeffekt: Mer kommunikasjon og informasjonsdeling på tvers av landegrenser

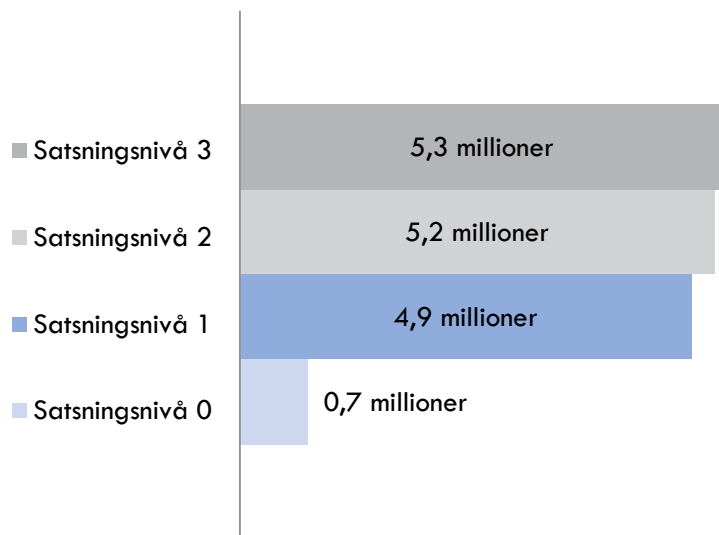


Nytte- og kostnadsvirkninger av en satsing på MT@EC

Øvrige ikke-prissatte virkninger	
Mer standardisering av språk	Oversettelsene bygger i større grad på et felles corpus.
Lavere språklig kvalitet	Feil eller uegnet ordvalg eller grammatiske feil i oversettelsen.
Flere misforståelser og mer støy i kommunikasjonen på tvers av landegrensler	På grunn av ukritisk bruk eller mangelfull etterredigering.
Bedre ivaretagelse av personvern og samfunnssikkerhet	Større sikkerhet/bedre håndtering av sensitive opplysninger enn ved bruk av gratis oversettelsesverktøy.
Tilgjengeliggjøring av informasjon og dokumenter på nynorsk	Mer tekst blir tilgjengeliggjort på nynorsk.
Konkurransen i markedet for maskinoversettelse	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 konkurransevidende effekter.

Antall segmenter i de ulike satsingsalternativene

Estimert antall segmenter som innhentes i en 10-års periode



Antagelser

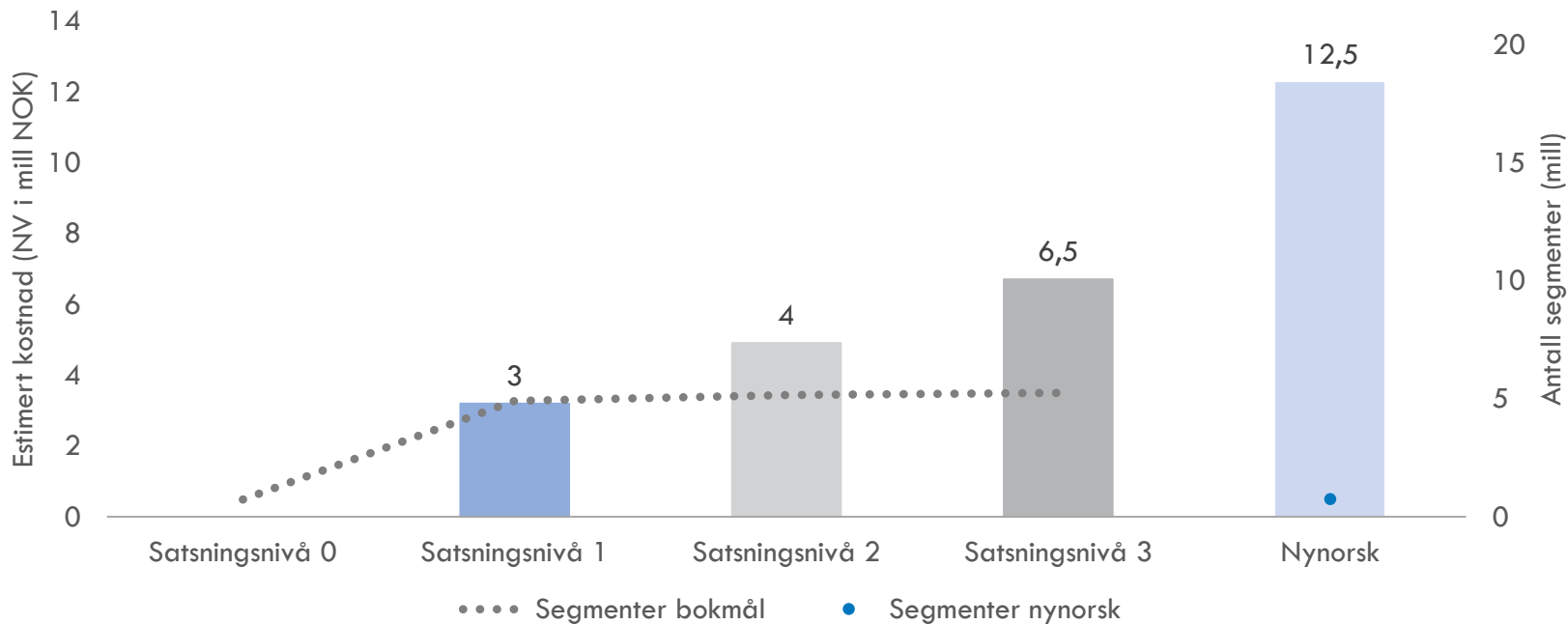
- Det finnes 733.081 segmenter i dagens corpus
- Ved å etablere en rutine for årlig avlevering kan 300.000 segmenter samles inn årlig
- Omtrent 1,2 millioner segmenter vil kunne innhentes fra minnet til profesjonelle oversettere
- Omtrent 250.000 segmenter vil kunne innhentes dersom man åpner for å benytte parallelle tekster i word-dokumenter eller lignende
- 100.000 segmenter oversettes til MT@EC i satsningsnivå 3

Vurdering av virkninger ved ulike satsingsnivå

Virkning	Konsekvens	Konsekvens	Konsekvens	Nynorsk
	Nivå 1	Nivå 2	Nivå 3	
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/+
Konkurransen i markedet for maskinoversettelse	0/-	0/-	0/-	0/-

Estimerte kostnader ved de ulike satsingsnivåene

Totale kostnader vs. antall segmenter



Forutsetninger for kostnadsestimering

Satsningsnivå 1	Satsningsnivå 2	Satsningsnivå 3	Nynorsk
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)	-
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	-
	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	-
		100 000 segmenter produseres, 9 ord per segment i snitt og 2 kr per ord	700.000 segmenter produseres, 9 ord per segment i snitt og 2 kr per ord

Usikkerhetsmomenter i vår analyse

Omfang av tekst som kan benyttes i corpus

Kvaliteten på maskinoversettelsene som oppnås innenfor de ulike fagområdene

I hvilken grad vil verktøyet tas i bruk?

Samlet vurdering og anbefaling

- 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.
- Satsingsnivå 1 utelukker ikke muligheten for en videre satsing i fremtiden:
 - 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.
- En høy kostnad, kombinert med et lavt rapportert behov for nynorskoversettelser gjør at vi vurderer at det ikke er grunnlag for å anbefale en satsing på nynorsk i MT@EC.

Vedlegg

Utvalget av statlige virksomheter i kartleggingen

Gjennomførte dybdeintervjuer

1. Utenriksdepartementet:
 - Avdeling for EØS-oversettelser
 - Seksjon for oversettelser
 - Europaavdelingen v/Geir Bekkevoold
2. Statens vegvesen
3. Forbrukerrådet
4. Mattilsynet
5. Statistisk Sentralbyrå
6. Sjøfartsdirektoratet
7. Justisdepartementet
8. Forbruker Europa
9. Barne- og likestillingsdepartementet
10. Domstolsadministrasjonen
11. Stortinget

Svar på spørreskjema mottatt fra

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. Helsedirektoratet
13. Helse- og omsorgsdepartementet (HOD)
14. Integrerings- og Mangfoldsdirektoratet (IMDi)
15. Innovasjon Norge
16. Klima- og miljødepartementet (KLD)
17. Kripes
18. Kunnskapsdepartementet (KD)
19. Lovdata
20. Miljødirektoratet
21. NAV
22. Nasjonal kommunikasjonsmyndighet (NKOM)
23. NOKUT
24. Nærings- og fiskeridepartementet
25. Olje- og energidepartementet (OED)
26. Petroleumstilsynet (Ptil)
27. Riksrevisjonen
28. Samferdselsdepartementet (SD)
29. Skattedirektoratet
30. Utlendingsdirektoratet (UDI)

Spørreundersøkelse blant ansatte

Spørreundersøkelsen sendt ut til:

1. Mattilsynet (19/22)
2. BLD (6/17)
3. Innovasjon Norge (3/18)
4. Forbruker Europa (2/3)
5. Statens vegvesen (2/25)
6. Sjøfartsdirektoratet (4/4)
7. Finanstilsynet (4/4)
8. UDI (6/9)

Eksisterende og fremtidige anvendelser i EUs online tjenester/plattformer

EU-webportaler hvor MT@EC er implementert i dag

IMI	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.
SOLVIT	Portal for privatpersoner, som skal bistå med å problemer knyttet til brudd på EØS-rettigheter.
Eur-Lex/ nLex	Portal som samler alle medlemslandenes nasjonale lovverk.
CircaBC	Kommunikasjon og informasjon ressurscenter for administrasjon, bedrifter og innbyggere.
TED	Portal for kunngjøring av offentlige innkjøp i EU.

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
Open Data Portal	Felles europeisk åpen portal for tilgang til åpne datainfrastrukturer
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

De fleste interne oversetterne benytter allerede digitale hjelpemidler

Virksomhet

Digitale hjelpemidler som benyttes til oversettelse (CAT-tool)

UD - EØS-oversettelse	SDL Trados Studio
UD - Seksjon for oversettelse	MultiTrans Prism
Statens vegvesen	Memsource
Sjøfartsdir.	SDL Trados Studio
Finanstilsynet	<i>Benytter ingen verktøy</i>
Stortinget	<i>Benytter ingen verktøy</i>

- De fleste interne oversetterne i staten benytter et såkalt «CAT-tool» - «Computer-assisted translation-tool»
- Dette er et dataprogram som støtter og fasiliteter oversetterens arbeid.
- **Viktige funksjoner:**
 - Segmenterer teksten i setninger og presenterer segmentene på en egnet form.
 - Lagrer segmentene i et minne.
 - Søkbar minne og/eller automatiske forslag til oversettelse av segmenter basert på minne (100 prosent «match») og andre forslag basert på minne)

Foreløpig oversettes det ikke til norsk i EU-plattformer hvor MT@EC er integrert

- BLD/Forbruker Europa arbeider med å gjøre ODR tilgjengelig for norske forbrukere med oversettelse til/fra norsk ved hjelp av MT@EC.
- Norske forbrukere vil dra nytte av systemet når Norge kobler seg opp (forhåpentligvis tidlig 2017)

Foreløpig oversettes det ikke til norsk i øvrige EU-plattformer hvor MT@EC er integrert.

Innspill fra virksomhetene:

- Av webportalene hvor MT@EC er implementert benytter flere av virksomhetene CircaBC og Eurlex
- I tillegg nevnes en rekke databaser, webportaler og nettsteder som er mer spisset mot virksomhetenes fagområder.
- Brønnøysundregistrene peker på at det er behov for oversettelse til/fra norsk i IML. De tror det kan bidra til at det blir delt flere tekster og dokumenter mellom statlige virksomheter i EU ved hjelp av IML.
- Lovdata ser flere bruksmuligheter for MT@EC og tror det kan bidra til å gjøre norsk regelverk tilgjengelig på andre språk enn norsk (i nLEX eller på andre områder)