

Les i morgen:



Hege Duckert
intervjuer Alice Hoffman.

søndag:
KUNST

mandag:
BOK

tirsdag:
MUSIKK

torsdag:
FILM

Babelsk forvirring

Skal man tro Bibelen, så snakket alle mennesker det samme språket da de bygde Babels tårn.

Babels tårn skulle strekke seg helt til Guds himmelrike. For å forhindre at menneskene kunne kommunisere mens de bygde seg

oppover mot skyene, skapte Gud mange ulike språk, og den babelske forvirringen var komplett. Gud oppnådde det han ønsket, mens menneskene mislyktes med tårnet sitt. Og grekerne snakker fremdeles helt gresk for oss.



TIL HIMMELRIKE:
Babels tårn.

Ingen er bedre på å **oversette språk** enn mennesker. Men nå begynner maskinene å hale innpå.

Babels maskin

Babelfisken må være alle oversetteres våte drøm. En liten, gul fisk som du stikker inn i øret og umiddelbart forstår du alt som blir sagt til deg, på alle språk. Det var Douglas Adams som diktet opp dette vidunderet i «The Hitchhiker's Guide to the Galaxy».

DEN BABELSKE FISKEN eksisterer dessverre ikke, men drømmen om en universaldekode for alle verdens språk har eksistert lenge før Douglas Adams sendte Arthur Dent ut på en odysse i verdensrommet.

På 1950-tallet trodde forskerne at det skulle ta tre-fire år før maskinoversettelser skulle ta jobben fra translatørene. Men oppgaven har vist seg å være mye mer komplisert. Språk er den mest komplekse menneskelige atferden. Bare tenk på hvor frustrerende det kan være når ett fremmed ord, eksempelvis det franske ordet *couler*, har sju forskjellige betydninger, blant annet «synke» og «flyte».

Det er svært krevende – om ikke umulig, vil mange hevde – å skulle automatisere språkforståelsen, og i snart 60 år har dataingeniører jobbet med å få maskiner til å oversette som mennesker. Men det siste tiåret har utviklingen skutt fart, blant annet på grunn av 11. september og den mangedoblede datakapasiteten – det som kalles Moores lov.

Etter 11. september ble det ekstremt viktig for amerikanske myndigheter å kunne forstå hva al-Qaida og andre terrorister planla å gjøre via e-poster, nettsider og telefonsamtaler. Amerikanske myndigheter har investert store summer i utvikling av språkteknologi rettet blant annet mot arabisk, pashto og dari (som snakkes i deler av Irak, Afghanistan og Pakistan). Men også den stadig mer globale økonomien har økt behovet for raskt å kunne oversette kinesiske eller russiske forretningsdokumenter.

EN DEL AV DE nye språkredskapene får vi nytte gratis, selv om mye er forbeholdt etterretning og kommersielle selskaper.

På Internett har babelfisken en navnebror som spytter ut oversettelser på få sekunder. På Yahoo-eide <http://babelfish.altavista.com/> kan du oversette ord og tekstutdrag eller hele nettsider. Den er langt fra den beste tjenesten som finnes, men den gir en viss pekepinn på hva blant andre sørkoreanske nettaviser skriver om. Heldigvis har konkurransen om de beste maskinoversettelsene blitt skjerpet siden Google meldte seg på i

kampen. Men det er verdt å merke seg at resultatet blir mye bedre mellom to europeiske språk sammenliknet med et asiatiske og et europeisk.

SLIK SER ET tekstutdrag fra den franske avisen Le Monde ut når vi oversetter det fra fransk til engelsk med Google Translate: «*The draft budget 2008, presented by George W. Bush Monday, envisages a rise of 10,7% of the expenditure related to safety.*»

Fra arabisk til engelsk blir det mer kryptisk, som i dette utdraget fra en artikkel i den syriske nettavisen Furat: «*The Arab Republic of Egypt has been discussed of fighting in the Palestinian territories and how to move to stop this fighting.*»

Ikke særlig overbevisende resultat, men det er bedre enn ingenting og gir oss et innblikk i hva artikkelen handler om. En må likevel dra på smilebåndet over setningsoppbygningen, og i kjølvannet av Babel Fish har humoristiske nettsider oppstått, såkalte «rundtur-oversettelser». Her blir en tekst oversatt fra eksempelvis engelsk til tysk, fra tysk til fransk og tilbake til engelsk igjen. Resultatet er babelsk moro!

– Vi kan ikke regne med å få et maskinoversettelsessystem som oversetter slik mennesker gjør i overskuelig framtid. Rett og slett fordi et slikt system måtte inneha svært mange av de kunnskapene og evnene som mennesker har, sier Jan Tore Lønning, professor i informatikk ved Universitetet i Oslo.

Han har selv vært med på å utvikle en ny arkitektur for maskinoversettelser mellom norsk og engelsk i det som kalles LOGON-prosjektet (se sidesak). Og han mener maskinoversettelser har en viktig rolle å spille.

– Menneskelige oversettelser er dyre. Maskinene kan være et hjelpemiddel for å gjøre menneskelige oversettelser mer effektive, samt at de kan være et reelt alternativ der lavere kvalitet er akseptabelt, sier Lønning.

TIDLIGERE HAR maskinoversettelser vært basert på noe som heter regelbaserte systemer – altså et språks grammatikk og syntaks – men i løpet av 90-tallet begynte forskerne å eksperimentere med noe som heter parallelle tekster. Det vil si at en bruker store mengder av tekster på et språk, eksempelvis nyhetsmeldinger fra Reuters, og ut fra denne tekstmengden analyserer maskinen sannsynligheten for at ord og fraser i et språk skal ende opp som bestemte ord eller fraser i et annet språk, også kalt statistisk basert mas-

kinoversettelse. Systran (regelbasert) og Language Weaver (statistisk) er blant de største kommersielle aktørene innenfor maskinoversettelser.

I New York sitter seks dataingeniører og programmerer en maskinoversettelse som skiller seg ut fra Google og de fleste andre. Meaningful Machines heter selskapet, og hemmeligheten deres er en smart algoritme og en gigantisk ordbok.

– Vi ønsker å bringe verden tettere sammen og å gi folk verktøy til å kommunisere. Vi har en sterk følelse av at det vi jobber med er viktig for samfunnet og for kampen mot terror.

De vidløftige ordene kommer fra Mark Steinbaum, administrerende direktør i Meaningful Machines. Han leder en bedrift som har utviklet maskinoversettelser som gjør det svært bra i offisielle tester. Ideen bak de meningsfulle maskinene er at systemet bruker en gigantisk mengde med tekst fra språket teksten skal oversettes til (f.eks. engelsk), litt tekst fra språket teksten oversettes fra (spansk) og en enorm tospråklig ordbok. Meaningful Machines har bygd opp en spanske engelsk ordbok med om lag to millioner oppslagord, 20 ganger flere enn en standard Merriam-Webster-ordbok. Resultatet er dette, som teknologimagasinet Wired refererte til i en artikkel:

«*We declare our responsibility for what happened in Madrid just two and a half years after the attacks on New York and Washington.*»

IMPONERENDE. Hva er så ulempen med denne teknologien, for noe må jo det være siden den ennå ikke er på markedet etter seks års programmering?

– Vi må få ned tiden på oversettelsene. Kvaliteten er bra, men det tar for langt tid, forteller Steinbaum, som sier at det tar tre sekunder å oversette en setning med åtte til ti ord. Deretter må de utvide med flere språk enn spansk, samt skaffe flere investorer.

Det kan være at Meaningful Machine er inne på noe her, noe virkelig stort. Men la oss tenke oss at det en gang blir mulig å oversette alle språk med en dings, en meningsfull maskin. Hva skjer da?

Douglas Adams har nok en gang et svar. Ved å fjerne alle barrierer for kommunikasjon mellom forskjellige raser og kulturer, har babelfisken forårsaket de blodigste krigen i hele galaksens historie. Det hører også med til historien at babelfisken er det endelige beviset på at Gud ikke finnes.



Ideer

Tekst: **Bente Kalsnes**
ideer@dagbladet.no

«Vi ønsker å bringe verden tettere sammen og å gi folk verktøy til å kommunisere. Vi har en sterk følelse av at det vi jobber med er viktig for samfunnet og for kampen mot terror».

Mark Steinbaum,
administrerende direktør i
Meaningful Machines

Europeisk babbel

23 offisielle språk. Tre offisielle alfa-beter.

EU-systemet vet alt om språklig mangfold eller kaos, alt etter som hvordan man ser på det. De vet også at menneskelige over-

settere er dyre og langtekkelige. 1123 millioner euro (2005) eller 1 prosent av budsjettet koster det EU å ha 23 offisielle språk. 4900 mennesker jobber med oversettelser i EU-systemet.

– Vi kan ikke basere oss ensidig på maskinoversettelser, men vi bruker flere digitale verktøy i arbeidet, sier Luka Tomasi, kommunikasjonskoordinator i Kommissjonens direktorat for oversettelser.



SPRÅKKAOS: Det er 23 offisielle språk i EU.



Hva gjør denne fisken i øret mitt? Den oversetter for deg. Det er en babelfisk.

Douglas Adams, «The Hitchhiker's Guide to the Galaxy»



BABELFISKEN: I Douglas Adams «The Hitchhiker's Guide to the Galaxy» presenteres Babelfisken som gjør alle verdens språk forståelige. Det var kanskje noe som dette han så for seg? Fotomontasje: Lars Eivind Bones



MASKINOVERSETTELSE: Professor Jan Tore Lønning og 11 andre forskere har utviklet oversettelsesystemet LOGON.

Norsk språkmaskin

De ville ikke finne seg i at bare engelsk og andre store språk skulle maskinoversettes.

Informatikkprofessor Jan Tore Lønning ved Universitetet i Oslo har sammen med 11 andre forskere fra tre norske universiteter utviklet LOGON, en maskinoversettelse fra norsk til engelsk. Det var ikke et mål for gruppen å bygge et fullstendig oversettelsessystem, men de har laget en ny arkitektur for maskinoversettelser som gir mer velformede setninger, noe som ikke er tilfelle for alle oversettelsessystemer.

– Hva er de største problemene dere har møtt på i dette arbeidet?

– Med vår metode har det største problemet vært at vi ikke er i stand til å analysere og oversette alle setninger vi møter i en tekst. Det andre store problemet er flertydigheter. Når setningene blir lengre, øker også tallet på flertydigheter, sier Lønning.

LOGON ble støttet med 20 millioner kroner av Norsk forskningsråd mellom 2002 og 2006, og hensikten var å få norsk språkteknologi opp på et akseptabelt nivå for ikke å henge etter i utviklingen. Resultatet skal ikke kommersialiseres eller være en oversettelseside på Internett, men arbeidet vil kunne bygges videre på, forteller Lønning.

– Hvorfor er det så viktig å kunne oversette ved hjelp av en maskin?

– Behovet for oversettelser og for å forstå hverandre på tvers av forskjellige kulturer har økt de seinere årene som en følge av globaliseringen og økt reisevirksomhet. I Norge får vi ofte inntrykk av at behovet ikke er så stort fordi nesten alle kan engelsk. Men ønsker vi å bevare norsk som et bruksspråk på alle områder, er det viktig at vi kan oversette fra norsk til engelsk.

Kort fortalt har LOGON kombinert to metoder i deres maskinoversettelse – transfer eller overføringsmetoden, samt statistisk basert maskinoversettelse. Den første metoden sikrer at meningen blir overført og at setningene blir velformet, den andre metoden sikrer at den riktige oversettelsen blir valgt.