
DATORN SOM ÖVERSÄTTARE

Det finns en dröm om att datorer ska kunna fungera som översättare mellan vad som i datavetenskapliga kretsar kallas "naturliga språk". Alla som använt automatiserade översättningstjänster som **Google translate** inser att dessa system har en lång väg kvar innan de kan kallas bra. **Lars Ahrenberg** påpekar att hur vi bedömer systemen beror på vad vi använder dem till. Kanske ska vi inte fråga oss varför systemen fungerar så dåligt – när man tittar under ytan på de enkla algoritmer som används är det snarare förvånande att de fungerar så pass bra som de trots allt gör.

Text: Lars Ahrenberg

Intresset för att automatisera översättning har funnits nästan lika länge som datavetenskapen själv och många specialister med olika bakgrunder har känt sig kallade: lingvister av olika schatteringar, översättare, matematiker, ingenjörer, och även en och annan AI-forskare. Tendensen har dock länge varit i riktning mot mer AI och i synnerhet maskininläring. Detta gäller nu inte bara för översättning utan för språkteknik generellt. Där man förut såg en grupp lingvister som central för att formulera den kunskap ett språktekniskt system behöver, har intresset förskjutits mot att hitta så mycket relevanta texter som möjligt och sedan tillämpa mer eller mindre generella maskininlärningsmetoder på dessa texter.

Översättning kan ses som en beslutsprocess både för människor och maskiner. Förutsättningarna är dock olika. Översättningsprogram har i stort sett endast språkliga data till sitt förfogande, något slags lexikon som relaterar ord och fraser från två språk till varandra (en "översättningsmodell") och någon slags kunskap om vad som är en bra ordsekvens på målspråket (en "målspråksmodell"). Den kunskap om texters uppbyggnad, genrer, ämnen, syften och målgrupper som en yrkesöversättare har saknas hos maskinen.

En annan skillnad är att maskinens beslutsgrund oftast är reducerad till vad den kan hitta i en given mening. System som kan utnyttja hela texten när det översätter en mening finns på forskarnas agenda, men har ännu inte nått sådana resultat att den extra processorkraft som krävs är mödan värd. En följd av denna begränsning är förstås också att maskinen inte har några förutsättningar att ändra meningsuppdelingen i texten (annat än genom felbeslut).

När beslutsprocesser implementeras på maskiner tillämpas ofta någon form av sannolikhetsberäkningar. Den valda översättningen av en mening ska ha högre sannolikhet än

alla andra alternativ och sannolikheterna bestäms utifrån en viktning av de använda modellerna.

En effekt av användningen av maskininläring är att det idag går mycket snabbt att skapa ett system, givet en utvecklingsplattform och en för ändamålet tillräckligt stor mängd av parallella texter. Flera sådana plattformar finns att få tillgång till, även om de kräver datavetenskaplig kompetens för att kunna användas. Tillgång till parallella texter kan vara ett större problem, men den akademiska forskningen har fritt kunnat använda sig av olika EU-texter, exempelvis protokoll från EU-parlamentet. De stora IT-jättarna har byggt upp egna resurser, framför allt för engelska och andra stora språk.

Maskinens beslutsgrund är oftast reducerad till vad den kan hitta i en given mening

Genom att bygga parallella textresurser med engelska som ett av språken blir det också möjligt att bygga system mellan godtyckliga språkpar. Vill man ha ett system som översätter mellan, säg, 50 språk i godtycklig riktning krävs det 2450 (50 x 49) parallella textmängder om det ska göras direkt, men bara 49 om det görs med ett av språken som brygga (pivotspråk).

Det är mer tveksamt om utvecklingen också inneburit en höjd kvalitet. En svårighet om man vill försöka besvara den frågan är förstås vad kvalitet är för en automatisk översättning och hur den kan mätas. Ytterst måste kvalitet sättas i relation till användning, men forskningen har gärna fokuserat på textgrundade mått. Detta kan innebära att räkna översättningsfel eller göra bedömningar av hur

bra översättningen är. En viktig kvalitetsaspekt är givetvis översättningens överensstämmelse med källtexten, dess semantiska kvalitet, en annan den målspråkliga kvaliteten: hur bra den fungerar som en text på målspråket. Ytterst är man hänvisad till mänskliga bedömare för att kvantifiera kvalitet, men för att slippa den kostnad och tidsåtgång som detta kräver varje gång man vill mäta har man också utvecklat automatiska mått. Dessa baseras vanligen på graden av överensstämmelse mellan en översättning och en eller flera referensöversättningar: översättningar som gjorts av översättare, eller åtminstone människor. Ett viktigt krav på ett automatiskt mått är att det korrelerar med mänskliga bedömningar. Det har dock visat sig att texttyp, språkpar och testmaterial spelar stor roll för vilket mått som har störst korrelation, varför någon entydig vinnare inte kunnat utses. Ett annat problem är att man behöver använda de automatiska måtten inte bara för sluttestning utan också för optimering av systemens parametrar och då har de mer avancerade måtten en nackdel, eftersom de ofta kräver någon form av syntaktisk/semantisk analys av översättningen innan den jämförs med referenstexterna.

Det låter självklart att säga att en översättning är bättre än en annan från samma källtext om den har färre fel eller bättre semantisk överensstämmelse. Men den intressanta frågan är snarare om skillnaden är relevant. För den vanligaste användningen av maskinöversättning, det vill säga för dechiffring av en text på ett främmande språk, visar det sig att de flesta människor kan stå ut med en ganska låg målspråklig kvalitet och att semantiska konstigheter ofta kan övervinnas med hjälp av sunt förnuft.

Låt oss anta att vår förståelse av danska är begränsad och att vi använder oss av något populärt översättningssystem på nätet för att få mer detaljer om ett kommande oväder över Danmark vi hört rykten om. Vi hittar en text och tillämnar systemet och får ut följande på svenska:

DMI räknar nu Danmark under söndag kväll kommer att drabbas av en vanlig skyfall, som kan falla upp till 25 mm regn på bara en halvtimme.

De första danskarna att byta paraplyer med paraplyer kommer Schleswigers. Det är i södra Danmark, regn och åska flyga in i landet. Redan under söndags kan komma regn städer, men det är en början. 21:00 att storma tvättar över Danmark. På den tiden, den första regn drabbade södra Jylland och södra Fyn.

Den första meningen är fullt begriplig och ger oss nog så mycket information vi behöver, trots att den är gjord via engelska. Möjligen undrar vi varför ett sådant kraftiga skyfall beskrivs som vanligt, men här har systemet översatt originalets "et regulært skybrud" till det engelska "a regular cloudburst" och misslyckats med översättningen av "regulär". Den andra meningen är besynnerligare och visar förstås systemets bristande förmåga att tolka meningar som helheter, att förstå poängen med dem. Orsaken här är att engelskan använder "umbrella" om både paraplyer och parasoller så att originalets "De første danskere, der skal skifte parasoller ud med paraplyer bliver sønderjyderne" på engelska blivit "The first Danes to change umbrellas out with umbrellas will Schleswigers", vilket sedan resulterat i en lika misslyckad svensk mening. Som läsare kan vi ju dock helt

enkelt ignorera den och hämta fler begripliga detaljer om händelsen från resten av texten.

Från ett språkligt perspektiv ter det sig ju rätt oförnuftigt att översätta danska till svenska via engelska. Men det är ur ett globalt företagsperspektiv mest en fråga om resurser och vad som lönar sig.

Översättning som snabb lästjänst kan sägas vara maskinöversättningens egen unika nisch tillsammans med talöversättning via mobilappar. Men i sammanhang där översättare vanligen verkar är kraven på kvalitet betydligt högre. Många har därför frågat sig hur översättningsprocesser kan effektiviseras genom att utföras av människa och maskin i samverkan. En linje, numera ofta förkortad PEMT ("Post-Edited Machine Translation"), frågar sig hur ett redigeringsverktyg för maskinöversättningar bör utformas för att texten så snabbt som möjligt ska kunna nå erforderlig standard. Det finns de som hävdar att en översättare kan höja sin produktivitet dubbelt eller mer genom att övergå till PEMT. Det finns givetvis också kritiker som hävdar att

Ytterst måste kvalitet sättas i relation till användning

kvaliteten blir sämre bland annat därför att arbetet förlorar sitt innehåll för en utbildad översättare, varpå förespråkarna svarar att postredigering kan utgöra en egen specialitet. Klart är dock att högkvalitativ översättning fortfarande kräver den mänskliga intelligensen och överblicken.

Trots att mycket resurser redan satts in i forskning och utveckling kan man spekulera över vad en riktig kraftsamling skulle kunna åstadkomma. TAUS, en tankesmedja med säte i Nederländerna och inriktad på lokalisering och översättning globalt, har lanserat idén om ett *Human Language Project* i analogi med *The Human Genome Project*. TAUS vill överkomma språkbarriärer genom "a global collaboration between business, government, academia and individuals with the goal of making language data and technology accessible to all stakeholders in the world" och finansieras bland annat av EU:s program Horizon 2020.

Man kan fråga sig om inte en satsning lik den IBM gjorde med sitt Watsonsystem skulle kunna lyfta kvaliteten på maskinöversättningar. **Watson**, som kunde besegra mästare i *Jeopardy*, sågs med rätta som en triumf för AI-tekniker. Sant är att Watson utgör ett ordentligt genombrott på det delområde av språkteknologin som på svenska brukar kallas frågebesvarande system. IBM:s forskare gjorde en ingående analys av vad som krävs av en skicklig *Jeopardy*-spelare. Hen måste

1. förstå beskrivningar av typen "This country borders on Sweden, Norway, and Russia" och tolka dem som en fråga
2. associera sådana beskrivningar med någon företeelse,

i detta fall Finland

3. bedöma om associationen är rätt, eller i alla fall om den "känns" rätt

4. göra associationen och bedömningen av dess korrekthet snabbare än sina medtävlare

5. formulera responser i form av frågor som "What is Finland?"

Watson besitter således en hög grad av såväl språkförmåga, begreppskunskap och associationsförmåga. Om vi på samma sätt försöker beskriva en översättares förmågor (och då bortser från punkten 4) ser vi dock snabbt att dessa går avsevärt utöver Watsons. Även om beskrivningarna i *Jeopardy* kan vara komplexa och hämtade från snart sagt vilket ämnesområde som helst, så följer de en given mall, något som i än högre grad gäller responsen. Översättaren måste kunna förstå texter med många andra typer av konstruktioner (punkt 1) och dessutom associera dem med lämpliga motsvarigheter i ett annat språk (punkt 2). För varje associerad översättning har hon också en uppfattning om den

Översättning som snabb
lästjänst kan sägas vara
maskinöversättningens egen
unika nisch

känns rätt (punkt 3), men kan inte som Jeopardyspelaren avstå om den känns fel, utan är tvungen att söka sig till andra källor, det må vara speciallexikon, termdatabaser, kolleger eller ämnesexperter för att komma till ett beslut. Responsen slutligen (punkt 5), innebär att sätta ihop de associerade översättningarna till något som inte bara är en sekvens av grammatiskt korrekta meningar utan också utgör en bra text anpassad till syfte och målgrupp.

Resultatet av en sådan tänkt storsatsning är därför svårt att sia om. Dagens bästa system är dock onekligen fortfarande dumma snarare än intelligenta, deras kapacitet är kopplad till storleken på databaserna och till förmågan att rätt uppskatta sannolikheten för ett ord eller en fras utifrån information om en mycket begränsad lokal ordkontext. Att detta fungerar så bra som det ändå gör är en upptäckt av språktekniken, något ganska oväntat som språkvetenskapen borde ge oss en förklaring till.



Lars Ahrenberg är professor emeritus i datorlingvistik vid Linköpings universitet.