

# Matnyttig minikonferens om maskinöversättning

Den 27 november ägde SFÖ:s första minikonferens per video rum; ämnet var maskinöversättning (MT) och deltagarna – ca 50 stycken – fick avnjuta två intressanta presentationer från Translator Scandinavia plus lite diskussion. Här följer ett referat. (En [videoinspelning](#) finns också.)

Konferensen inleddes med en presentation av företaget Translator Scandinavia av dess vice vd Carolina Kühnemann – en presentation som framför allt vinklades mot MT.

Translator Scandinavia bildades 1993 av Gunnar Carlsson, som de som har varit med en tid i SFÖ säkert har träffat. Numera är det ett mellanstort företag i branschen och är medlem i förutom SFÖ även Språkföretagen, GALA, LocWorld, NTIF, Eulogia och ITI.

Företagets användning av MT tog fart år 2017, när den neurala MT-tekniken hade kommit i gång. Allt fler kunder hänvisade till MT, och många började skicka maskinöversatta texter för efterredigering. Det visade sig också att MT-tekniken allt mer avspeglade sig i krav på lägre priser, vilket i sin tur förstas tvingade Translator Scandinavia att effektivisera arbetet. Och det gjorde att man började undersöka de möjligheter som MT erbjuder. Det verktyg som man då framför allt siktade in sig på var Memsources, vilket fortfarande är den huvudsakliga metod man använder.

Att använda MT i större skala var dock inte något man gjorde i en handvändning. Här fanns flera stöttestenar, som i första hand gäller relationerna med leverantörerna/frilansöversättarna (eller "lingvister"); en tilltalande term som Carolina ofta använde):

- Hur mäter man produktiviteten så att prissättningen blir rätt (mer om detta nedan)? Att som med minnesbaserade översättningar använda ordpris vore ju det enklaste, men det fungerade inte.
- Hur skulle man på lämpligt sätt diskutera om den ändrade inriktningen med frilansarna? Det gällde ju att beskriva det hela på ett korrekt sätt men samtidigt undvika att i onödan skrämja bort folk. (Många frilansöversättare kunde antas ha en på förhand negativ inställning.)
- Samtidigt ville man förstås behålla sina leverantörer och hjälpa dem att utvecklas som efterredigerare. Dessutom ville man hitta nya, vilket inte var alldeles lätt eftersom det gällde att få tag i dem som både var villiga att jobba med efterredigering och även var bra på det.

Med hjälp av en konsult kom man fram till denna sjustegsmetod:

1. Sätt upp mål för övergången – varför ville man gå in för MT och efterredigering? Svaret var att man behövde göra det för att hänga med.
2. Var är man i dag?
3. Börja långsamt – men börja nu! Vänta inte tills "allting" är på plats, för det lär aldrig hända.
4. Välj generiska (allmäninriktade, icke-specialiserade) MT-motorer (tjänster för maskinöversättning) – avstå från egen "träning" av motorer. Carolina nämnde Googles och Microsofts betaltjänster samt DeepL (den senare dock inte när nordiska språk är inblandade). De generiska NMT motorerna ger tillräckligt bra resultat i de flesta fall och vi har en mycket bred och fragmenterad kundbas – att skapa specialiserade motorer är därför inte värt jobbet.
5. Se till att engagera leverantörer på ett sätt som är tillfredsställande för båda parter. Använd direktkontakt!
6. Utveckla ett hållbart arbetssätt – nämligen praktiskt, framför allt vad gäller logistiken.
7. Följ upp leveranser och prestanda. Detta är svårt och tar tid! Det gäller bland annat att ta reda på vilka som verkligen har nytta av MT.

Så till ett av de största och mest diskuterade problemen med MT, nämligen ersättningsmodellerna. Många läsare känner säkert igen följande tre alternativ:

- Betalning per timme. Det är enkelt och bra som startpunkt. Dessutom kan denna metod göra att man med tiden kan komma fram till ett ungefärligt ordpris (nästa alternativ) vilket har sina fördelar. En nackdel är förstås att man inte i förväg kan veta beloppet.
- Erfarenheten visar att en översättare normalt översätter X ord/timme, vilket ger ett ordpris vilket i sin tur indikerar ersättningen redan innan arbetet påbörjas (som när man jobbar med översättningsminne). Men det gäller att man har hamnat rätt, annars kan kvaliteten bli lidande (om översättaren känner sig tvungen att jobba extra snabbt för att få rimlig betalning).
- Redigeringsavstånd (editing distance). Fördelar: enkelt att beräkna med hjälp av CAT-verktyg. En nackdel är att man precis som med timpris inte vet ersättningen förrän jobbet är klart.

Hur är då läget på MT-fronten i dag för Translator Scandinavia? Jo, man använder MT för omkring 60-70 procent av alla jobb. Ibland kommer kunden med en egen MT-översättning, ibland använder man MT ändå därför att man tror att det kommer att spara mycket tid. Vad gäller sekretessen svarar Carolina på min senare mejl-fråga: "Vi jobbar inte med gratis motorer utan med betalda tjänster där sekretessen är garanterad. På samma sätt som tex. Microsoft garanterar sekretessen kring de dokument man laddar upp på Office365."

Men som antyds ovan är det en stor utmaning att följa upp översättningarnas kvalitet. Även administrationen när det gäller MT-uppdrag ställer extra krav. Så allt är inte en dans på rosor.

Efteråt undrade någon om det gick att skönja några trender när det gäller vad kunder respektive översättare föredrar (MT eller ej). Svaret var att eftersom detta är så pass nytt är det svårt att säga. Om kunden är en annan, stor byrå får man fundera på om man ska ta risken (nämligen att ersättningen inte motsvarar vad översättaren bör få betalt) eller ej. Med direktkunder kan problemet vara orealistiska förväntningar på prisreducering.

Inte heller vad gäller fördelningen i förväg MT-översatta texter respektive oöversatta texter där PEMT efterfrågas går det att säga något bestämt, vare sig om nuläget eller trend.

Vad gäller relationen till frilansarna är timdebitering ett bra alternativ; då kan man komma överens om ett cirka-timpris som eventuellt kan justeras efteråt. Metoden med redigeringsavstånd accepteras ofta om än inte entusiastiskt.

Någon undrade över den algoritm som Memsources använder för mätning av redigeringsavstånd. Det är lite krångligt, fick vi veta, även om Memsources har förklarat. Det liknar ju rätt mycket "vanligt" ordpris (med "Trados-trappa") vilket känns lite hemmastakt, men metoden passar inte alla texttyper.

Det påpekades också att redigeringsavståndet noll (ingen ändring av MT-förslaget) kan innebära att översättaren har grunnat en hel del innan hen beslutar sig för att godta förslaget – då är ju ersättningen noll kronor inte acceptabel. Carolina höll med och påpekade att deras system är konstruerat så att översättaren får betalt även om hen inte ändrar MT-träffen. Det betyder att en 100-procentig träff (redigeringsavstånd noll) betalas med 20-25 procent av det reducerade ordpriset.

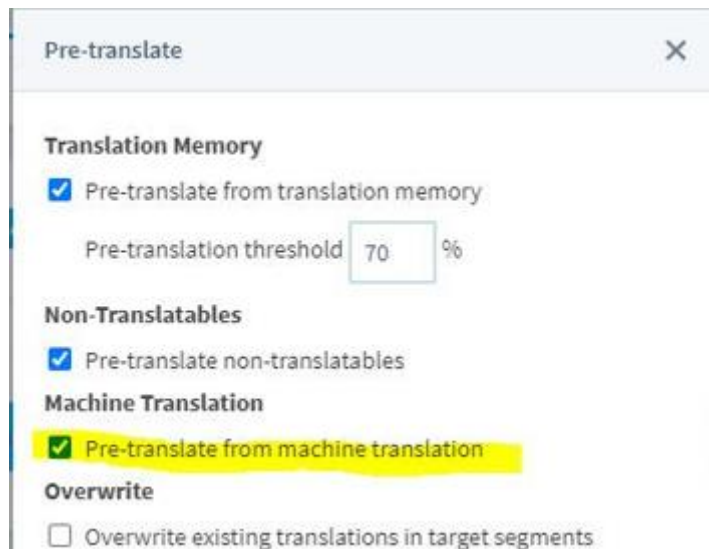
Avslutningsvis poängterade Carolina att arbetet med MT innebär stora utmaningar: det innebär ett annorlunda sätt att jobba än med översättningsminne – vilket också kunden måste fås att förstå. (Och det är olika för olika översättare). Det handlar också om hur mycket tid kunden vill betala för – och en kort tid kan medföra ett högt timpris!

Efter Carolina kom Anne-Marie Silvester, som berättade om själva praktiken. Hon kommer ursprungligen från Storbritannien men har varit i Sverige i drygt 20 år och har jobbat hos Translator Scandinavia nästan lika länge – som projektledare, granskare och översättare; sedan 2007 som enbart översättare och granskare. Hon var den första på företaget som började använda MT, för ungefär ett år sedan, och hon tycker att det fungerar bra; hon gör nu nästan alla jobb i MT. Metoden är särskilt bra för myndighetstexter, lättare tekniska texter, viss affärskommunikation. Det finns också vissa ämnesområden som lämpar sig särskilt väl, till exempel utvecklingssamarbete. Mindre bra fungerar det med reklamtexter/copy, men det börjar bli bättre även där.

Anne-Marie tror att det är en fördel att ha jobbat mycket med granskning av andras texter; det ger ett sätt att tänka som passar bra för MT.

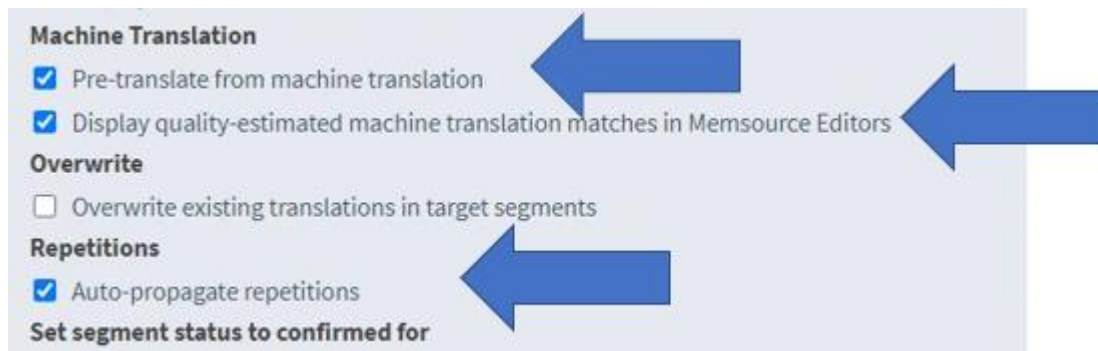
Vad är det då som är bra? Jo, MT spar tid och tangentnedslag, särskilt när man blir van. Man slipper tänka ut somliga formuleringar och leta fram somliga termer. Men det beror förstås mycket på textens typ.

Därefter gick Anne-Marie igenom arbetsgången i Memsourc. (Som man kan tänka sig jobbar Anne-Marie på ungefär samma sätt när hon granskar andra översättares resultat.) Moment 1 är projektledarens inställningar:



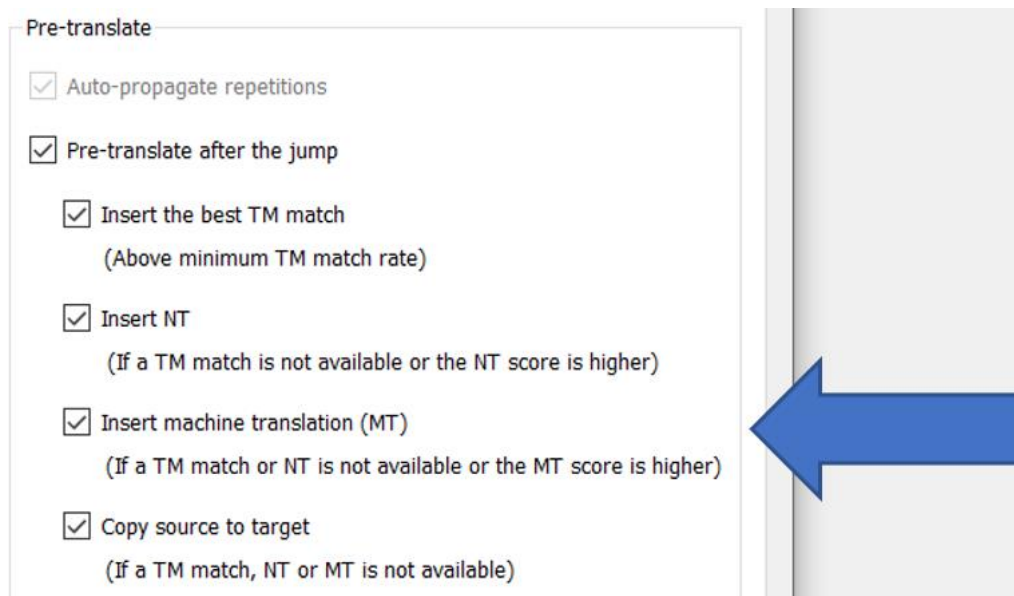
Inställningen av tröskelvärdet beror på kvalitetsförhållandena TM-MT och kan vara lite känsligt; ett för högt värde här kan innebära att MT-förslagen får ”orättvist” försteg. Och funktionen föröversättning (dvs. att MT-förslaget automatiskt förs in som målsegment i stället för att bara visas som separat förslag) har sina för- och nackdelar. Den spar tid, men vissa översättare gillar den inte – det kan leda till att man godtar förslagen lite för lätt.

Fler projektledarinställningar:



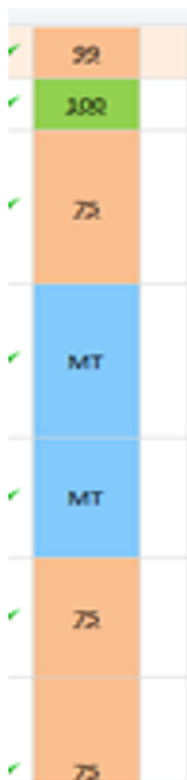
Även automatisk ifyllning av upprepningar kan ha sina sidor; kan vara praktiskt men man kan behöva se upp.

Så till översättarens egna inställningar:



”MT score” är en beräkning av kvaliteten hos MT-förslaget enligt Memsources egna algoritmer (mer om det nedan). Givetvis kan det vara så att tröskelvärde för TM-träffar och denna ”score” tillsammans medför att man får ett MT-förslag inlagt automatiskt samtidigt som kanske motsvarande TM-träff är bättre. Då får man kolla, vilket kan vara besvärligt.

Så kommer vi till kvalitetsbedömningen av MT-träffarna i Memsources. Det här är ”skalan”:



- 100%: Perfect MT output—probably no post-editing required
  - 99%: Near-perfect MT output—possibly minor post-editing required, mostly for formatting or punctuation issues
  - 75%: Good MT match, but likely to require some human post-editing
  - No score (blue): This means MTQE cannot confidently identify the quality (it may be high or low), so the output needs to be checked by a Linguist.
- (from the Memsources website)

Till vänster ser vi ett utsnitt ur hur det ser ut i Memsources. Man får ta poängen med en nypa salt, men funktionen kan spara tid när man har vant sig vid den.

Och så här kan gränssnittet i Memsources se ut:

	Filter source text	Filter target text			
	Source: sv	Target: en-gb			
1	1.	1.	x	100	
2	Helsingborg-(Centralstationen)--Höganäs-(Gästhamnen)-sträckans-längd:	Helsingborg-(Central-Station)--Höganäs-(Guest-Harbour)-Length-of-the-route:	x	75	⌵
3	22,2-km	22.2-km	x	100	⌵
4	Se-höjdkurva-längst-ner.	See-height-curve-at-the-bottom.	x	MT	⌵
5	Från-Helsingborg-cyklar-du-norrut,-passerar-Sofiero-slott-och-de-långa,-härliga-stränderna-vid-Domsten-och-Viken.	From-Helsingborg-you-cycle-north,-passing-Sofiero-Castle-and-the-long,-beautiful-beaches-at-Domsten-and-Viken.	x	75	
6	Leden-går-längs-vattnet,-och-utsikten-över-ett-glitrande-Öresund-är-minst-sagt-magisk.	The-trail-runs-along-the-water,-and-the-view-of-a-glittering-Sound-is-magical-to-say-the-least.	x	75	
7	Kattegattleden-börjar--eller-slutar--i-en-av-Nordens-äldsta-städer,-Helsingborg.	The-Kattegat-trail-begins--or-ends--in-one-of-the-Nordic-region's-oldest-cities,-Helsingborg.	x	75	
8	Den-första-sträckan-tar-dig-från-stadskärnan-till-Gröningen-och-Fria-bad,-två-populära-badplatser-som-lockar-med-glass,-sol-och-bad.	The-first-stretch-takes-you-from-the-city-center-to-Gröningen-and-Fria-bad,-two-popular-bathing-spots-that-attract-with-ice-cream,-sun-and-swimming.	x	75	
9	Sitt-ner-och-njut-ett-tag,-för-snart-kommer-ledens-första-uppförsbacke.	Sit-down-and-enjoy-for-a-while,-because-soon-the-trail's-first-uphill-will-come.	x	MT	
10	Lite-längre-fram-ligger-Sofiero-slott-och-1 Sofiero-Gård 1, med-gårdsbutik-och-café,-som-båda-är-väl-värda-ett-besök.	A-little-further-ahead-are-Sofiero-Castle-and-1 Sofiero-Farm 1,-with-a-farm-shop-and-café,-both-of-which-are-well-worth-a-visit	x	75	

Här kan man först titta på TM-översättningarna – men i detta fall var TM tomt:



Anne-Marie brukar annars börja med att sortera (eller "filtrera" med andra CAT-verktygs term) fram de enklaste, mest mekaniska översättningarna först, nämligen till att börja med de vanligaste repetitionerna:

	Source: sv	Target: en-gb		
2	Helsingborg {Centralstationen}—Höganäs {Gästhamnen}-sträckans-längd:	Helsingborg {Central-Station}—Höganäs {Guest-Harbour}-Length-of-the-route:	x	75
16	Helsingborg {Centralstationen}—Höganäs {Gästhamnen}-sträckans-längd:	Helsingborg {Central-Station}—Höganäs {Guest-Harbour}-Length-of-the-route:	x	75
38	Helsingborg {Centralstationen}—Höganäs {Gästhamnen}-sträckans-längd:	Helsingborg {Central-Station}—Höganäs {Guest-Harbour}-Length-of-the-route:	x	75
3	22,2-km	22.2-km	x	100
17	22,2-km	22.2-km	x	100
39	22,2-km	22.2-km	x	100
4	Se-höjdkurva-längst-ner.	See-height-curve-at-the-bottom.	x	MT
18	Se-höjdkurva-längst-ner.	See-height-curve-at-the-bottom.	x	MT
40	Se-höjdkurva-längst-ner.	See-height-curve-at-the-bottom.	x	MT
1	1.	1.	x	100
5	Från-Helsingborg-cyklar-du-norrut,-passerar-Sofiero-slott-och-de-långa,-härliga-stränderna-vid-Domsten-och-Viken.	From-Helsingborg-you-cycle-north,-passing-Sofiero-Castle-and-the-long,-beautiful-beaches-at-Domsten-and-Viken.	x	75

Sedan kanske segment som inte ska översättas:

	Source: sv	Target: en-gb
1	1.	1.
61	Anders-Ebefeldt	Anders-Ebefeldt

Anne-Marie brukar också lägga in rätt många kommentarer för senare åtgärder.

Det kan noteras att Memsource inte kan hämta in översättningar från termdatabaser. Och att egennamn ofta "översätts" är ett typiskt drag. Det gäller att hela tiden vara uppmärksam!

I den slutliga kvalitetskontrollen bör man särskilt ge akt på oredigerade målsegment –

12	Unedited NT/MT fuzzy match
13	Unedited NT/MT fuzzy match
14	Unedited NT/MT fuzzy match
15	Unedited NT/MT fuzzy match

– liksom frågetecken för terminologi och stavning:

#	Type	
42	Terminology	missir Katte
58	Terminology	missir Katte
8	Spelling	cente
12	Spelling	defen
13	Spelling	harbc
14	Spelling	Kulla
21	Spelling	cozy

Andra problem kan gälla ordföljd, falska vänner, namn, titlar och tempus samt aktiva/passiva verbformer.

Sammanfattningsvis kan man säga till tveksamma frilansöversättare att branschen redan kräver att man kan jobba med MT. Tåget har gått! Även om man kan konstatera att det ibland går för fort:

- Käll: Vi har ett högt fokus på till exempel tekniska lösningar
- MT: We have a high focus on, for example, technical solutions
- Efter PE: We're focusing sharply on aspects such as technical solutions
- SV: Ett exempel är murarna av fjärsagnejs.
- EN: One example is the walls of remote snow gnomes.

Nedtecknat av Mats Dannewitz Linder (som drog igång minikonferenserna 2007, höll i dem till och med 2013 och är glad över att de fortfarande hålls).