

Taligenkänning på SVT

Anna Zettersten, Avdelningschef SVT Undertext

Bakgrund

I SVT:s sändningstillstånd finns tydliga krav på att uppdraget omfattar alla människor, exempelvis barn och unga, personer med etnisk minoritetsbakgrund och personer med funktionsnedsättning. Tillgänglighet är en högt prioriterad fråga för SVT. Företaget söker ta hänsyn till att varje individ har olika förutsättningar att ta till sig utbudet utifrån skiftande intressen och behov. Anpassning av utbudet görs för att bland annat överbygga funktionshinder.

Enligt det krav som avser tillgänglighet av tv-sändningar för personer med funktionsnedsättning ska SVT uppfylla vissa kvoter för de olika tillgänglighetstjänsterna. Den tekniska utvecklingen, inte minst avseende IT, gör att nya sätt att ta del av tillgänglighetstjänster växer fram. För att kunna utöka textningen i högre grad krävs nya tekniska lösningar för textproduktion vid direktsända program.

Taligenkänningsteknik förutsätter att en språklig infrastruktur finns tillgänglig. Inom ramen för pilotprojektet har viss omvärldsanalys gjorts och projektet har tagit del av erfarenheter om hur taligenkänningsteknik används och fungerar på flera andra håll i Europa, bl.a. Belgien (för flamländska), Danmark och Storbritannien. Dessa länder har haft tillgång till språkdata-baser på de egna språken sedan flera år tillbaka.

SVT har aktivt deltagit i pilotprojektet som rör utvecklingen av en svensk taldata-bas och taligenkänningsteknik för textning av TV-program. Metoden bygger på en så kallad "skuggtalare" som lyssnar vad som sägs och talar in det som blir den text som visas i bild.

Sammanfattningsvis vill SVT framhålla att erfarenheterna från pilotprojektet varit mycket goda då tekniken kunnat leverera inom ramen för de specifika kvalitetskrav som formulerats gällande snabbhet och korrekthet.

Pilotprojektet

För publiken är skillnaderna mot nuvarande teknik (högstastighetstangentbordet Velotype) främst att

1. när taligenkänningsmotorn korrekt igenkänner ett ord och återger det i skrift så blir det korrekt språkligt, medan när den misstolkar så blir felet väldigt stort. Med tangentbord kan det bli fler antal fel, men fler av dem är begripliga även om de inte är korrekt språkliga.
2. den sammantagna fördröjningen av texten i utsändning blir troligen oförändrad då ytterligare en station tillkommer i sändningskedjan.

Vid direktsändningsarbete textar man 10-15 minuter, sedan tar en kollega vid 10-15 minuter och så byts man av tills passet är slut. Metoden är densamma oavsett teknik (tangentbord eller taligenkänning). På SVT har vi dagliga direktsändningar som är 3 timmar eller längre. De längre är indelade i flera pass för olika team. Man textar i max 4 timmar, sedan byts teamet mot ett nytt. I Belgien och Storbritannien använder man samma metod, men textar via taligenkänning och då i max 2,5 timmar.

Den kognitiva processen är densamma oavsett metod – direkttextning kräver redigering i huvudet simultant med textproduktionen. Den viktigaste och mest svårrekryterade kompetensen för direkttextning är inte att kunna skriva på tangentbordet eller använda taligenkänningstekniken, utan just denna simultanförmåga.

Det anses allmänt (baserat på erfarenhet och forskning) i Europa och andra delar av världen att ca 3 timmar reelltidstextning är vad man maximalt mäktar med under ett arbetspass. Därmed kan att byte från tangentbords- till taligenkänningsteknik inte per automatik öka mängden producerad text för publiken. (Och i själva verket textar den enskilda direkttextaren på SVT redan större mängd än direkttextare på många andra PS-bolag i Europa.) Däremot är utbildningstiden för taligenkänning kortare än för tangentbord. Det går snabbare att få direkttextare som lär sig taligenkänning i produktion än tangentbordstextare.

Erfarenheterna från andra marknader visar att övergången från tangentbord till taligenkänning sker gradvis. Såväl i Belgien som Storbritannien och Danmark har man börjat med några få taligenkänningstextare och några projekt och sedan utökat vartefter, en process som tagit flera år.

För att fungera optimalt behöver språkdatan kontinuerligt fyllas på och uppdateras vilket innebär förändrade arbetssätt och utökad förberedelsestid inför sändningar. En annan skillnad mot tangentbordsarbete är att taligenkänning kräver avskildhet för att inte störa andra i närheten, vilket kan innebära flytt och/eller ombyggnation.

Vi ser också att man förutom direkttextning kan använda taligenkänningsmotorn och språkdatan till andra arbetsuppgifter inom media, bl.a. transkribering. När taligenkänning används för direkttextning till publik måste direkttextaren ange skiljetecken ("punkt", "utropstecken" etc) vilket inte är viktigt vid textinsamling för internt bruk.

Vi tror att det är lämpligt att fortsätta utveckla tekniken under 2017 och skulle gärna påbörja reella textningsprojekt som visas för publiken. Test behöver även göras för att undersöka vilka andra typer av programproduktioner som bäst lämpar sig för taligenkänning (sport, politik, nyheter, underhållning, barnprogram etc), eftersom att endast använda sig av tekniken för textning av vädersändningar inte är kostnadseffektivt i en produktionsmiljö. Vi behöver också reda ut närmare hur kompetensprofilen skiljer sig från nuvarande samt kunna komma med inspel till hur en utbildning skulle kunna se ut.