



## Utvärdering av SVT:s prototyp för skuggtalarteknik

---

## Sammanfattning

År 2014 fick Post- och telestyrelsen (PTS) ett regeringsuppdrag som bland annat inkluderade att ta fram en begränsad prototyp av ett taligenkänningsystem som kan användas för direkttextning av tv-program på Sveriges television (SVT). Utvecklingstiden för prototypen startade i november 2015 och avslutades i december 2016.

### Stelacons uppdrag

PTS har bett konsultföretaget Stelacon att utvärdera om den nya skuggtalartekniken är tillräckligt bra för att användas som direkttextningsmetod. Utvärderingen ska utifrån ett användarperspektiv svara på hur egenskaper såsom fördröjning och träffsäkerhet påverkas vid användning av skuggtalarteknik jämfört med den tidigare Velotype-tekniken. Utvärderingen ska också undersöka vad effekterna blir hos tv-bolagen av att byta direkttextningssystem, dvs. vad skillnaden blir i arbetsresurser, licenskostnader, kompetens hos direkttextningspersonalen, arbetsmiljö och dylikt.

### Studiens resultat

Studiens resultat tyder på att tittarens upplevelse av direkttextning inte kommer att förändras avsevärt om SVT väljer att implementera skuggtalartekniken. Den skillnaden som tittaren kan märka vid byte till skuggtalarteknik är att texten presenteras i ett annat flöde och rytm än tidigare samt att felen i texten blir annorlunda. Möjligtvis kan även felen i texten minska.

Användartesterna indikerar även att textens fördröjning är den aspekt som har störst påverkan på användbarheten av undertexterna, detta resultat är oberoende av vilken målgrupp testpersonen tillhör. För många målgrupper är fördröjningen avgörande för om testpersonen använder undertexten över huvud taget. Givet behovet av minskad fördröjning bland alla tittargrupper så kan det eventuellt finnas ett behov av att utvärdera till vilken målgrupp undertextning vänder sig till. Om målgruppen anses vara bredare än personer som har grav hörselnedsättning eller dövhet så bör ett övervägande göras om det finns möjlighet att göra någonting för att minska fördröjningen, exempelvis senarelägga livesändningar eller förbereda mer material för att minska fördröjningen. För personer som använder texten som stöd till ljudet så är det viktigt att ljud och text kommer i nära anslutning till varandra annars har de ingen användning av texten.

Den tydligaste positiva effekten av ett införande av skuggtalartekniken är att utbildningen av texterna är betydligt kortare än vid användning av dagens teknik. Detta medför framförallt en ökad flexibilitet för tv-bolag som snabbt kan ta in nya resurser i organisationen vid resursbrist eller semestrar samt eventuellt minska lönekostnader. Omvärldsanalysen har även visat att det också ger möjlighet att möta de ökande kraven på andel textade program.

---

## Innehållsförteckning

<b>1. Inledning.....</b>	<b>4</b>
1.1. Bakgrund .....	4
1.2. Syfte med studien .....	5
1.3. Metod för genomförandet.....	5
1.4. Definitioner .....	6
1.5. Disposition.....	6
<b>2. Utveckling av skuggtalarteknik .....</b>	<b>8</b>
2.1. Utveckling av skuggtalarteknik i Sverige .....	8
2.2. Utvecklingen av skuggtalarteknik i andra länder .....	10
2.2.1. Anledningen till att övergå till Skuggtalarteknik.....	10
2.2.2. Skillnader i undertextning beroende på kultur och språk .....	10
2.2.3. Egna lösningar .....	11
<b>3. Tittarnas upplevelse av prototypen .....</b>	<b>13</b>
3.1. Genomförande av användartester.....	13
3.2. Urval av målgrupper och filmklipp.....	14
3.3. Resultat från användartesterna .....	16
<b>4. Effekterna för tv-bolag .....</b>	<b>19</b>
<b>5. Slutsatser .....</b>	<b>21</b>

---

# 1. Inledning

## 1.1. Bakgrund

Sveriges Television (SVT) skall i enlighet med nu gällande sändningstillstånd tillse att minst 65 procent av de direktsända programmen i företagets huvudkanaler (SVT1 och SVT2) är textade<sup>1</sup>. Textning av tv-program sker för att personer med hörselnedsättning skall kunna ta del av SVT:s programutbud. Det beräknas att drygt 1 miljoner svenskar<sup>2</sup> har hörselskador eller hörselnedsättning av sådant slag att textning är nödvändigt för att uppnå full begriplighet.

Frågan om programtextning, både för direkt- och förberedande textning i tv, har i olika omgångar uppmärksammats av riksdagen och riksdagens kulturutskott. Ett flertal motioner med krav på bättre textservice har behandlats och textningen av SVT:s program utgör en prioriterad fråga för den nu sittande statliga public service-utredningen.

En teknik som används ibland annat Storbritannien är så kallad skuggtalarteknik som är ett talarberoende system för textning av direktsända tv-program. En professionell inläsare talar in den text som önskas användas som undertext. Då systemet dels baseras på den talades röst och dels att den talande anpassar röstläge och röstform till taligenkänningssystemet kan önskvärd precision och hastighet säkerställas. Förutom att skuggtalaren talar in det som sägs i tv-programmet sammanfattar talaren texten kontinuerligt så att det endast är det viktigaste som sänds ut i undertexterna. Eftersom människan kan tala fler ord i minuten än vad som kan skrivas, har denna teknik potential att vara mer effektiv än tidigare tekniker som har används för att producera undertexter till direktsända tv-program.

Den nuvarande metoden för direkttextning i Sverige utvecklades på 1980-talet och bygger på snabbinskrivning av text, via ett Velotype-tangentbord. Velotype-tekniken har begränsningar bland annat när det gäller precision och tillgång till personer som behärskar tekniken. Därutöver sker inte någon utveckling längre av de tekniska hjälpmedel som används. Således räcker inte denna teknik för att tillgodose de allt hårdare undertextkrav som ställs på tv-bolagen. I några länder har därför en övergång skett till taligenkänning för hantering av direkttextning. Den nya tekniken har visat sig vara en bra metod. Man använder då i första hand så kallade "skuggtalare" som formulerar och uttalar den text som skall visas. Automatisk taligenkänning förutsätter emellertid att en språklig infrastruktur finns tillgänglig. Av naturliga skäl har skapandet av denna grundval främst inriktats på ett begränsat antal världsspråk. Taligenkänningstekniken är därför idag enbart förbehållen programföretag som verkar inom stora språkområden.

År 2014 fick myndigheten Post- och telestyrelsen (PTS) ett regeringsuppdrag att utreda förutsättningarna för utveckling av en nationell språkresursbank samt att ta fram en prototyp av en sådan språkresursbank (se regeringsbeslut N2014/2840/ITP). Regeringsuppdraget inbegriper även framtagande av en begränsad prototyp av ett taligenkänningssystem som kan användas för direkttextning av tv-program. PTS genomför uppdraget i nära dialog med Institutet för språk och folkminnen (ISOF) och Sveriges Television AB (SVT).

---

<sup>1</sup> Kulturdepartementet, *Krav på tillgänglighet till tv-sändningar för personer med funktionsnedsättning avseende Sveriges television AB och Sveriges utbildningsradio AB*, KU2013/2310/MFI

<sup>2</sup> HRF, Äh, det var inget viktigt om hörselskadades situation i Sverige, 2007

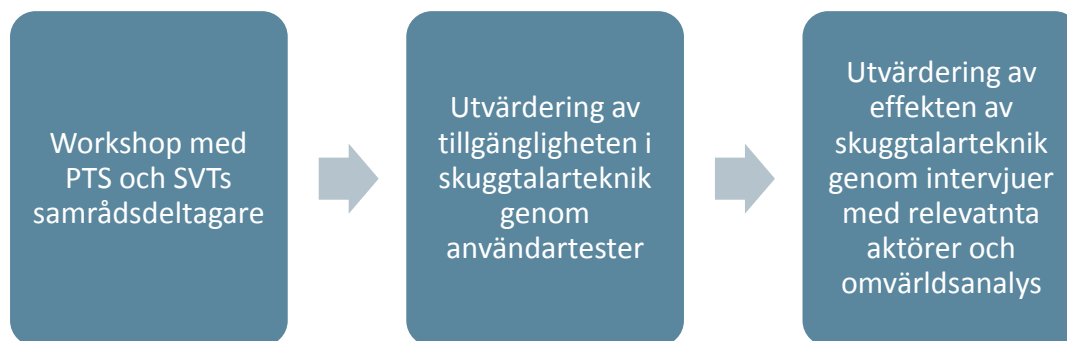
Under 2015 utlyste PTS, tillsammans med SVT, en offentlig upphandling gällande utveckling av en taligenkänningsprototyp som utnyttjar skuggtalarteknik för att direkttexta tv-program inom vädergenren. Det finska språkföretaget Lingsoft vann den offentliga upphandlingen, och har integrerat prototypen i SVT:s befintliga undertextsystem. Syftet med utvecklingen och integrationen var att utvärdera om tekniken är tillräckligt bra för att användas i SVT:s framtida tv-sändningar. Utvecklingstiden för projektet startade i november 2015 och avslutades i december 2016.

## 1.2. Syfte med studien

Inom ramen för regeringsuppdraget har PTS bett konsultföretaget Stelacon att utvärdera om den nya skuggtalartekniken är tillräckligt bra för att användas som direkttextningsmetod. Utvärderingen ska utifrån ett användarperspektiv svara på hur egenskaper såsom fördröjning och träffsäkerhet påverkas vid användning av skuggtalarteknik jämfört med den tidigare Velotype-tekniken. Utvärderingen ska också undersöka vad effekterna blir hos tv-bolagen av att byta direkttextningssystem, dvs. vad skillnaden blir i arbetsresurser, licenskostnader, kompetens hos direkttextningspersonalen, arbetsmiljö och dylikt. Utvärderingen avses ligga till grund för rapporteringen av regeringsuppdraget.

## 1.3. Metod för genomförandet

Stelacon har genomfört två delar inom ramarna för uppdraget. Dels utvärderat vad slutanvändarna anser om den nya tekniken och dels analyserat vad det innebär att implementera skuggtalarteknik utifrån ett tv-bolagsperspektiv. Processen har sett ut enligt följande:



Workshopen genomfördes tillsammans med SVT och PTS samrådsdeltagare och syftade till att ge samrådsdeltagarna möjlighet att diskutera utvärderingsfasen och komma med förslag på vad som bör inkluderas. Under workshopen kommenterades också vilka filmklipp som bör väljs ut i användartesterna samt diskuterat för- och nackdelar med skuggtalarteknik och tidigare teknik. Baserat på workshopen togs sex vädersändningar som textats med skuggtalarteknik fram till de kommande användartesterna.

Användartesterna syftade till undersöka hur målgruppen upplever direkttextningen från den taligenkänningsprototyp som Lingsoft har utvecklat och integrerat i SVT:s befintliga undertextsystem. Testpersonerna fick se vädersändningarna med direkttexter producerade med både skuggtalarteknik och Velotype-teknik – både med och utan ljud.

I slutet av varje filmklipp fick testdeltagarna uppge (mellan 5-0) hur välfungerande de tyckte att texterna var – 5 är perfekta undertexter, och 0 är underkända undertexter.

---

Dessa två moment (workshop och användartester) gav slutkonsumentperspektivet av den nya prototypen.

Effektutvärderingen syftade till att utreda vad effekterna blir av att använda skuggtalarteknik. Exempel på aspekter som belystes var hur arbetsresurser, licenskostnader, arbetsmiljö, kompetens- och utbildningskrav hos direkttextarpersonalen och liknande variabler påverkas av att använda skuggtalarteknik. För detta intervjuades SVT, TV4-gruppen, undertextningsleverantörer samt talteknologiexperter från den akademiska världen. Därutöver intervjuades även representanter från programföretag, och undertextningsleverantörer från andra länder där skuggtalarteknik implementerats för att kunna dra lärdomar om vilka effekter som tekniken gett där.

Intervjuerna har genomförts med semistrukturerade djupintervjuer som metod. Semistrukturerad djupintervjuteknik är en kvalitativ metod som syftar till ge intervjupersonen öppna frågor och låta denna tala fritt om ämnet inom ramarna för uppdraget.

#### 1.4. Definitioner

Nedan presenteras definitioner av de begrepp som används i rapporten.

**Skuggtalarteknik** – En metod som används för att texta direktsända tv-program där en professionell inläsare talar in den text som önskas användas som undertext.

**Språkresurser** – Tal- och textresurser som bland annat används för att bygga upp talbaserade tjänster.

**Språkresursbank** – En databas innehållande språklig data såsom text, tal och ljud i digital form. En databas som kan användas för att utveckla intelligenta språkteknologiska tjänster.

**Talarberoende taligenkänningsystem** – Taligenkänningsystem som är anpassat efter en viss individ, och får därmed bättre precision då systemet har anpassats efter talaren.

**Talaroberoende taligenkänningsystem** – Taligenkänningsystem som inte är anpassat efter en viss individ, utan kan användas av många olika människor.

**Taligenkänning** – Processen att elektroniskt eller datatekniskt tolka mänskligt, naturligt tal.

**Talresurser** – Inspelat tal som används som grundsten och resurs till talteknologi, såsom dialogsystem och dikteringssystem.

**Talsyntes** – Artificiellt tal som ska efterlikna människans röst skapat med datorer.

**Talteknologi** – Tal som används som grundsten och resurs till talteknologi, såsom skuggtalarteknik eller en personlig assistent

**Textresurser** – Text som används som grundsten och resurs till textteknologi, såsom lexikon och stavningskontroll.

**Velotype** – Ett ackordbaserat tangentbord som används för att texta tv-program.

#### 1.5. Disposition

I kapitel två beskrivs utvecklingen av skuggtalartekniken i Sverige och internationellt.

I kapitel tre presenteras Stelacons studie om tittarnas upplevelse av skuggtalarteknik som direkttextningsmetod. Tillvägagångssätt och resultat från användartester presenteras.

---

Effekterna vid en eventuell implementering av skuggtalarteknik beskrivs i kapitel fyra.

Studiens slutsatser presenteras i kapitel fem.

---

## 2. Utveckling av skuggtalarteknik

Historiskt har det funnits ett antal olika tekniker för att direkttexta tv-program, som alla varit tangentbordbaserade. Inledningsvis användes vanliga tangentbord vars styrka var mycket hög träffsäkerhet, men hastigheten nådde oftast endast upp i ca 90-100 ord/minut, vilket är långt ifrån talhastighet. Detta innebar att mängden information som förmedlades antingen var tvunget att reduceras markant eller så var det stora fördröjningar innan texten nådde rutan.

Sedan dess har två mer anpassade tangentbordslösningar utvecklats, den så kallade Velotype-tekniken (ca 140 ord/minut) samt ett stenograftangentbord som kan producera upp till 225 ord per minut. Båda dessa tekniker har hög träffsäkerhet och används fortfarande i många länder. Men, för båda dessa tangentbord krävs mycket lång utbildning (ca 1-1,5 år för Velotype-skrivbordet och 2-3 år för stenografiskrivbordet), vilket innebär stora begränsningar i utbud av personer som kan direkttexta tv-program i respektive land samt stora kostnader för utbildningskostnader.

Dessa begränsningar innebar att många länder började leta efter alternativa lösningar i mitten av 2000-talet.

### 2.1. Utveckling av skuggtalarteknik i Sverige

Tekniken att använda skuggtalarteknik i syfte att direkttexta tv-program är i viss mån beroende av att någon form av språkresursbank finns tillgänglig då det annars blir en mycket kostsam process att utveckla skuggtalartekniken. Det är därför naturligt att dessa två frågor har följt varandra i det svenska utredningsväsendet. Tankar på att i någon form nyttja taligenkänning för textning av tv-program har enligt SVT funnits sedan slutet av 90-talet. Även om teknik och metoder alltså föreligger för programtextning via automatisk taligenkänning har bristen på infrastruktur för svenska språket varit ett avgörande hinder för att ta i bruk en modern teknik. SVT tog under 2006 initiativ till en offentlig upphandling i syfte att låta utveckla en taligenkänningsmjukvara på svenska. Upphandlingen fick dock avbrytas eftersom ingen anbudslämnare kunde infria de uppställda kraven eller ansåg det vara av intresse att utveckla den nödvändiga grundstrukturen för ett så litet språk som det svenska.

Den prototyp som PTS och SVT har upphandlat inom ramarna för nuvarande regeringsuppdrag är en prototyp för skuggtalarteknik. I prototypstadiet valdes att endast använda tekniken på vädersändningar då detta är en avgränsad och relativt enkel språkbas som passar väl för en begränsad prototyp. Under slutet av 2016 var prototypen klar av leverantören Lingsoft och levererades till SVT. Med hjälp av prototypen har SVT tagit fram olika klipp där skuggtalartekniken används för direkttextning. Dessa klipp är endast framtagna för att användas vid användartesterna i denna studie och avspeglar inte hur tekniken kommer att se ut vid en eventuell framtida implementering. Exempelvis presenteras inte texterna på samma sätt som de skulle göra om tekniken implementeras i verkliga sändningar, istället placeras texten en bit nedanför bilden, inte i nederkant på bilden som den gör i dagens sändningar, se skärmdump med exempel i figur 1.



Figur 1. Bild från filmklipp med skuggtalarteknik som direkttextningsmetod.



En annan skillnad mellan de olika textningsteknikerna är hur texten presenteras i klippet och i vilken rytm som textningen sker. Idag använder sig SVT av texter i två rader som försvinner när den nedersta raden är full. I prototypen används istället rullande text där den nedre raden flyttas upp till den översta raden när den är full. Texten kan också komma ut i sjok och inte ord för ord, vilket är fallet vid textning med Velotype-teknik.

För att få en bild över dagens situation där Velotype-tangentbord används för att texta både livesända och förinspelade program har Stelacon analyserat sex väderklipp som textats med Velotype-teknik. I analysen undersöktes antal felaktiga ord, rättade ord samt genomsnittlig fördröjning. Dessa har sammanställs för respektive filmklipp, se tabell 1 nedan.

Tabell 1. Sammanställning av Stelacons analys av sex direkttextade filmklipp

Filmklipp	Teknik	Felaktiga ord	Andel ord som rättas	fördröjning
1	Velotype	3	67 %	5,2 sek
2	Velotype	3	100 %	5,5 sek
3	Velotype	1	100 %	5,4 sek
4	Velotype	4	50 %	5,3 sek
5	Velotype	2	50 %	5,5 sek
6	Velotype	5	20 %	5,1 sek

Den genomsnittliga fördröjningen räknades ut genom att var tionde sekund räkna hur många sekunders fördröjning det var på ordet som sas i klippet. Sedan räknades ett medeltal ut av de

---

olika fördröjningarna i klippet. Detta har gjorts av Stelacons personal. Alla klipp är så gott som samma längd.

Analysen av de olika filmklippen tyder på att den genomsnittliga fördröjningen varierar väldigt lite mellan de olika klippet. Alla klipp har en fördröjning på mellan 2,1 och 5,5 sekunder. I klippet som är två minuter är det mellan ett och fem felaktiga ord, andelen av dessa som rättas varierar mellan de olika filmklippen.

## **2.2. Utvecklingen av skuggtalarteknik i andra länder**

Skuggtalarteknik har används som direkttextningsmetod för tv-sändningar sedan år 2001. Först ut var Storbritannien där BBC har använt skuggtalarteknik för att direkttexta tv-sändningar sedan april 2001 då världsmästerskapen i snooker utspelades. Även om infrastrukturen fanns på plats fortsatte dock BBC ändå att använda tangentbordslösningar för direkttextning i mycket stor utsträckning upp till 2007-2008. Sedan dess har de dock mer och mer gått över till skuggtalarteknik vid textning av både direkt- och förinspelade program. Ungefär samtidigt som direkttextningen började användas mer i Storbritannien började även andra länder studera möjligheterna att använda tekniken för sin undertexter. I majoriteten av länderna (Storbritannien, Frankrike, Spanien, Tyskland, USA, Kanada, Nederländerna och Australien) är Ericsson det företag som levererar den tekniska lösningen till tv-bolagen. Därmed är den tekniska lösningen relativt snarlik i majoriteten av de länder där skuggtalarteknik finns tillgängligt. Vad som skiljer sig åt är snarare respektive lands språk, kultur och lagstiftning. Omvärldsanalysen kommer därför fokusera på ett antal faktorer som driver skillnader och visa vad dessa beror på med exempel från olika länder.

### **2.2.1. Anledningen till att övergå till Skuggtalarteknik**

Majoriteten av de tv-bolag som använder sig av skuggtalarteknik har valt metoden av ett antal samvarierande anledningar. Utbildningen för att kunna hantera både Velotype-lösningar och stenografilösningar är mycket långa och dyra, vilket både innebär att personalen blir dyrare samtidigt som tv-bolagen har svårare att arbeta i en flexibel organisation där det är enkelt att rekrytera ny personal vid behov. Tekniken ses också som relativt gammal och i exempelvis Storbritannien var det svårt att få personer att vilja utbilda sig inom stenografi, vilket innebar att den sista skolan stängde under 2005. Samtidigt ökade efterfrågan på mer textning från döva tittare i mitten av 2000-talet. Även tillsynsmyndigheterna ökade sina krav på mängden textade och direkttextade program. Dessa faktorer innebär att tv-bolagen hade svårt att rekrytera kompetent personal till deras dåvarande direkttextningsmetoder och de övergick därmed till skuggtalarteknik.

### **2.2.2. Skillnader i undertextning beroende på kultur och språk**

Eftersom att skuggtalarteknik förlitar sig på taligenkänningsteknik, där en skuggtalare talar in de ord som denne önskar få med i undertexterna, kan resultatet av skuggtalartekniken variera beroende på hur språket är uppbyggt. Detta var också en av anledningarna till att PTS fick regeringsuppdraget att undersöka hur väl tekniken fungerar med just det svenska språket. De länder som utnyttjar skuggtalarteknik för direkttextning och framkommit som mest betydande är beskrivna nedan.

I Storbritannien och Spanien strävar de nationella tv-bolagen mot att allt som sägs i sändningarna även ska återges i undertexterna, en så kallad ord-för-ord princip som innebär att stor mängd text produceras och principen påverkar även fördröjningen av undertexterna.

---

Anledningen till detta anses vara kulturellt betingad, där användarna är vana vid att få all information och troligtvis skulle uppleva att information gick förlorad om undertexterna började sammanfattas i större utsträckning. Då även fördröjningen av text har varit central i dessa länder har det inneburit att korrektheten i undertexterna inte nödvändigtvis måste uppnå 100 procent, utan snarare 96-98 procent.

I just Storbritannien har BCC fått utstå en del kritik för direkttextningskvalitén. Ofcom är den myndighet i Storbritannien som genomför tillsyn och granskning av direkttextningen i landet och som ett led i detta arbete har de, mellan 2013 och 2015, genomfört mätningar för att identifiera förbättringsområden. Granskningen genomfördes vid fyra olika mättillfällen och kvaliteten mättes bland annat i träffsäkerhet, fördröjning och undertexthastighet på de fem största brittiska tv-bolagen. Träffsäkerheten uppgick, i alla fyra testtillfällen, över den rekommenderade träffsäkerhetsnivån på 98 procent. Noterbart är dock att metoden för att mäta träffsäkerhet var i den så kallade NER-metoden, och inte WER<sup>3</sup> som används i Sverige, därav går det inte att dra några direkta jämförelser mellan träffsäkerhetsnivån. Fördröjningen varierade mellan 5,1 och 5,6 sekunder från att ett ord uttalades i tv-sändningen till att rätt ord dök upp i tv-rutan, vilket ligger över Ofcoms rekommendation på max 3 sekunders fördröjning. Vidare genomförde organisationen Action on Hearing Loss in the UK en landsomfattande undersökning, under 2013, i syfte att kartlägga användarnas uppfattning om direkttextningen i landet. Resultatet av undersökningen visade på att användarna anser att det största problemet med direkttextningen i nuvarande tappning är den stora fördröjningen, medan de näst största problemet anses vara inkorrekt språk.

Även Inom fransk och belgisk public service används uteslutande skuggtalarteknik som direkttextningsmetod. I och med franskans relativt komplexa språkuppbyggnad samt förväntningar om helt korrekta undertexter påverkar det bland annat fördröjningen i direkttexterna. Detta innebär att direkttexterna i Frankrike arbetar i par där den ena personen gör inläsningen, medan den andra personen korrigerar all text innan den går ut till tittarna. Systemet innebär att normaltillståndet för fördröjning är 10 sekunder, med vissa undantagsfall kan det gå upp emot 20 sekunder. I Belgien har tv-bolagen istället löst problemet med den språkliga utmaningen genom att låta direkttexterna sammanfatta innehållet i sändningarna i stor utsträckning istället för att läsa ord alla ord som sägs. Detta medför att fördröjningen inte är lika stor som i Frankrike, men många grupper som är beroende av undertexterna anser att allt för stor del av det som sägs försvinner i sammanfattningarna.

### 2.2.3. Egna lösningar

Ericssons lösning är i dagsläget framförallt anpassad efter de större språken, eller språk som är nära besläktade med de större språken, vilket medfört att ett antal länder tvingats utveckla egna lösningar.

#### 2.2.3.1. Tjeckien

I Tjeckien har de nationella tv-bolagen krav på sig att 70 procent av alla program ska textas med 98 procents träffsäkerhet. För detta används skuggtalarteknik som direkttextningsmetod i sportsändningar samt vissa direktsända program såsom "*Lets Dance*". I Tjeckien har även Public service-bolaget Czech Television använt sig av ett talarberoende system för att sända

---

<sup>3</sup> WER och NER är två olika metoder att beräkna träffsäkerheten i direkttextning. Båda metoderna utgår från ett antal parametrar såsom antal ord, antal fel samt antal korrigeringar. WER-metoden delar i större utsträckning upp felen i olika typer.

---

parlamentariska debatter i direktsänd tv. Vid dessa sändningar sitter en direkttextare och arbetar med formgivning, rättstavning, färgkodning och liknande redigeringsverktyg medan taligenkänningsystemet tolkade ljudkällorna efter bästa förmåga. Resultatet av detta är ca 90 procents träffsäkerhet<sup>4</sup>.

Utvecklingen av teknikplattformen för direkttextad skuggtalarteknik inleddes under 2006. Inledningsvis var planen att detta skulle göras tillsammans med Red Bee Media (nu ägt av Ericsson AB), men de tekniska lösningarna som Red Bee Media kunde erbjuda var allt för standardiserade och ansågs inte tillräckligt anpassade för det tjeckiska språket. Därför utvecklade Czech Television egna tekniska lösningar i samarbete med en lokal leverantör. En av lösningarna är att direkttextaren använder sig av ett tangentbord för skiljetecken, vilket enligt deras erfarenhet förenklar processen för direkttextaren. Vidare har direktsända program i Tjeckien en fördröjning med ca 3 sekunder, vilket innebär att skuggtalaren vinner extratid och fördröjningen av texten i rutan är därför begränsad till ett fåtal sekunder.

#### 2.2.3.2. Finland

I Finland används skuggtalarteknik som undertextningsmetod för förinspelade tv-sändningar. Public service-bolaget Yle använder sig av leverantören Lingsoft för att få undertextningen till tv-sändningarna. Då tiden inte är en avgörande faktor, vilket är fallet vid direkttextning, har skuggtalaren gott om tid på sig att tala in det denne vill få ut i undertexterna och har tid att justera eventuella felaktigheter.

Lingsoft menar att skuggtalartekniken innebär en produktivitetökning med cirka 20 procent jämfört med tidigare tekniker, vilket innebär att mindre tid behöver användas till respektive program och därmed även mindre resurser. Detta resonemang gäller dock endast vid förinspelade tv-sändningar. Lingsoft har även noterat att direkttextarnas ergonomiska situation har förbättrats i och med införandet av skuggtalarteknik. Detta då de olika teknikerna tillåter möjligheten att i större utsträckning variera sin arbetsställning, vilket har mindre påfrestning på axlar, armar och händer.

---

<sup>4</sup> I Tjeckien används den så kallade WER-metoden för att räkna träffsäkerhet

### 3. Tittarnas upplevelse av prototypen

Det finns flera målgrupper som har behov av att program textas. Vissa använder texten som stöd till talet medan andra är helt beroende av texten för att förstå programmets innehåll. Den direkttextningsmetod som används idag får kritik från funktionshindersorganisationer och användare som anser att kvalitén på texten är för låg med stor fördröjning, mycket fel och allt för stort utelämnande av viktig information. Under den inledande workshopen med funktionshindersorganisationer visade organisationerna stort intresse för den nya prototypen och förhoppningar om att tekniken kan innebära en förbättring för deras medlemmar.

För att utvärdera hur tittarna upplever undertexter som är framtagna med skuggtalarteknik genomförde Stelacon användartester på Lingsofts prototyp med personer med funktionsnedsättning, äldre, personer som inte har svenska som modersmål samt en referensgrupp med personer utan påvisad funktionsnedsättning.

#### 3.1. Genomförande av användartester

Användartester är en kvalitativ metod som syftar till att utvärdera användbarheten och ge en djupare förståelse för hur något kan utvecklas eller förtydligas för ökad användbarhet och tillgänglighet. Under användartesterna fick testpersonerna se sju olika filmklipp med textade vädersändningar. Det första filmklippet var textat med ett Velotype-tangentbord och användes som ett referensklipp till de resterande klippen, se bild från klippet med Velotype-teknik i figur 2. Testpersonen fick sedan se sex klipp där texten är skapad med skuggtalarteknik. Ordningen på klippen som är textade med skuggtalarteknik varierades mellan testtillfällena för att undvika att inläring och vana påverkade testresultatet.

Figur 2. Bild från filmklipp med Velotype-tangentbord som direkttextningsmetod.



Användartestet delades upp i tre delar, se figur 3 nedan. Inledningsvis fick testpersonen information om studien samt fick svara på frågor om tidigare upplevelser och synpunkter på direkttextning. Sedan fick testpersonen se på de olika filmklippen och svara på frågor om varje filmklipp. Testet avslutades med en intervju med testpersonen.

Figur 3. Användartestetets tre delar



### 3.2. Urval av målgrupper och filmklipp

Tillsammans med uppdragsgivaren PTS och SVT bestämdes vilka målgrupper som skulle delta i användartesterna samt vilka filmklipp som var relevanta att visa under testerna. I tabellen nedan presenteras de målgrupper som deltagit i studiens användartester.

Målgrupp	Antal (st)
Personer med lätt hörselnedsättning	2
Personer med måttlig hörselnedsättning	4
Personer med grav hörselnedsättning	5
Personer med dövhet	4
Personer som inte har svenska som modersmål	2
Personer som är äldre än 75 år	3
Personer utan påvisad funktionsnedsättning	2
<b>Totalt</b>	<b>22 st</b>

Personer med olika grad av hörselnedsättning genomförde användartester då graden av hörselnedsättning påverkar på vilket sätt undertexterna används. Personer med lätt eller måttlig hörselnedsättning lyssnar och använder texten som stöd medan personer med grav hörselnedsättning istället kan vara helt beroende av texten och lyssnar lite eller inte alls på talet. En referensgrupp med personer utan påvisad funktionsnedsättning ingick också i urvalet. Syftet med dessa tester var att ge en referenspunkt för att säkerställa om resultaten från de övriga testerna beror på respektive funktionsnedsättning eller om det är något som alla användare upplever. Testpersonerna rekryterades genom Stelacons kontakter hos funktionshindersorganisationer samt kontakter från tidigare genomförda studier.

Sju filmklipp valdes ut till användartesterna. Eftersom att parametrar såsom detaljåtergivning och fördröjning motverkar varandra, många rättningar leder exempelvis till en korrektare text

men också längre fördröjning, så är det svårt att uppnå perfekta undertexter. I figur 4 nedan visas exempel på hur olika parametrar påverkar varandra vid direkttextning.

Figur 4. Olika parametrars påverkan på varandra vid direkttextning



För att förstå vilka parametrar som tittarna anser är viktigast och på så sätt få kunskap om hur SVT bör arbeta och prioritera så har SVT tagit fram klipp med stor variation i träffsäkerhet, fördröjning, antal ord, detaljåtergivning och antal rättningar. Den stora variationen mellan klippen gör att testpersonernas preferenser enklare kan framgå under ett användartest.

I tabell 2 nedan presenteras egenskaperna för de olika filmklippen. Antal textade ord i klippen varierar mellan 128 och 275 stycken. Klippet som är textat med Velotype-teknik har betydligt färre ord än det genomsnittliga antalet ord i resten av klippen. Några av filmklippen har varierad fördröjning, ibland ligger ljud och text i fas och ibland ligger ljudet före texten. Förutom fördröjning och antal ord så presenteras även andra aspekter i klippen såsom meningslängd.

Tabell 2. Fördröjning och antal ord i filmklippen.

Filmklipp	Teknik	Antal ord	Fördröjning	Kommentar
1	Velotype	128	6,5 sek	Text producerat med Velotype-tangentbord
2	Skuggtalare	247	3,5 sek	Genomsnittligt klipp
3	Skuggtalare	207	7,8 sek	Lång fördröjning
4	Skuggtalare	275	2,4 sek	En del fel. Meteorologen talar snabbt
5	Skuggtalare	209	1,9 sek	Meteorologen talar snabbt
6	Skuggtalare	259	3,9 sek	Långa meningar
7	Skuggtalare	181	3,7 sek	Korta meningar

Sett till antal felaktiga ord i texten så är det fler felaktiga ord i skuggtalade texter jämfört med texter som produceras med Velotype-tangentbord. De felaktiga orden rättas i större utsträckning i textning som görs med Velotype-tangentbord än som produceras med skuggtalarteknik.

### 3.3. Resultat från användartesterna

Det finns brister vid textning av direktsända program idag, detta uppger flera testpersoner i den inledande intervjun. När testpersonerna inledningsvis fick se ett väderklipp som är textat med Velotype-tangentbord var det flera som ansåg att texten är rörig och att det är mycket av talet som inte textas. Flera testpersoner med hörselnedsättning uppger också att de i vanliga fall stänger av ljudet när de ser på direktsända program då det upplevs som störande när texten har en allt för stor fördröjning jämfört med talet.

Resultatet från användartesterna visar att det finns aspekter som påverkar tittarnas upplevelse av direkttextningen mer än andra. Den aspekt som tycks ha störst påverkan är textens fördröjning. Denna aspekt påverkar framförallt tittare som tar del av både ljud och text då skillnaden mellan tal och text blir tydlig. För personer som använder texten som stöd till talet, exempelvis personer som inte har svenska som modersmål, så är det extra viktigt att ord och text kommer i nära anslutning till varandra, vid för lång fördröjning fungerar inte texten längre som stöd utan blir snarare ett störningsmoment och tittaren väljer då ofta att stänga av texten.

*”Om det är så synkroniserat som möjligt så finns det möjlighet att få stöd av det som sägs. Men man behöver kanske inte ha all text utan istället nyckelorden.”* – Testperson med grav hörselnedsättning

Testpersoner som inte får stöd av ljudet när de ser på tv, exempelvis döva eller personer med grav hörselnedsättning, upplever inte fördröjning lika störande som andra grupper. Några hörselskadade testpersoner uppger dock att det är förvirrande när en allt för stor fördröjningen gör att texten handlar om vädret i Skåne och meteorologen pekar på Norrland. För dessa grupper är det istället avgörande att texten täcker in det som sägs i sändningen, så att de inte missar någon viktig information.

*”Det meteorologen säger blir helt fel när jag läser. Hade nog hängt med bättre med bara ett utav elementen men inte med både tal och text.”* – Testperson med hörselnedsättning

Testpersonerna fick i slutet av användartestet svara på hur de upplever rättningar av felaktiga eller felstavade ord i direkttextningen. De flesta testpersonerna uppgav då att de föredrar när det är få rättningar då allt för många rättningar ofta leder till ökad fördröjning. Ökad fördröjning upplevs som ett större problem än några felaktiga ord då felen ofta är av sådan karaktär att texten ändå är förståelig, enligt testpersonerna. Om felen är för omfattande eller att meningarna innehåller så pass mycket fel att det inte är möjligt att förstå innebörden måste dock rättningar ske, uppger testpersonerna. Äldre testpersoner anser, i större utsträckning än andra målgrupper, att det är viktigt att fel i texterna rättas då de är rädda att viktig information annars kan gå förlorad.

*”Få rättningar med många felaktiga ord. Bättre att få en jämn rytm för förståelsen än att all stavning ska vara perfekt. Dock får det inte vara helt fel ord.”* – Testperson med grav hörselnedsättning

Testpersonernas betygsättning av filmklipp pekar också på att fördröjningen är den avgörande faktorn. Klipp 5, med kort fördröjning och få fel, är det klipp som upplevs som bäst, se tabell 3. Filmklippet med längst fördröjning, klipp 3, var det klipp som fick lägst snittbetyg från testpersonerna. Klippet som textats med Velotype-teknik, klipp 1, får också ett lågt betyg av testpersonerna. Detta klipp har, trots den relativt låga



ordmängden, bland den längsta fördröjningen av alla de testade klipp. Stelacons omvärldsanlys tyder på att en så pass, relativt sett, kort fördröjning som uppvisats i skuggtalarklippen högst troligtvis inte är möjlig att uppnå vid en eventuell implementering av tekniken i skarpt läge. Intervjuer med aktörer i Europa visar på att den genomsnittliga fördröjningen vid användning av skuggtalarteknik är ca 5 sekunder. Även om fördröjningen sannolikt inte kommer att nå samma nivå som i användartesterna så visar resultaten tydligt att fördröjningen är den aspekt som påverkar tittarnas upplevelse av direkttextning mest.

Tabell 3. Snittbetyg för de olika filmklippen som användes i användartesterna.

Filmklipp	Teknik	Antal ord	Fördröjning	Snittbetyg
5	Skuggtalare	209	1,9 sek	3,8
7	Skuggtalare	181	3,7 sek	3,2
2	Skuggtalare	247	3,5 sek	3,2
6	Skuggtalare	259	3,9 sek	2,9
4	Skuggtalare	275	2,4 sek	2,5
1	Velotype	128	6,5 sek	2,4
3	Skuggtalare	207	7,8 sek	2,2

Det är inte några markanta skillnader i betygsättningen mellan de olika målgrupperna i studien. Döva är den målgrupp som gett ett något högre genomsnittsbetyg än andra målgrupper. Detta beror troligen på att fördröjningen inte är lika påtaglig för denna målgrupp som för andra målgrupper. Denna målgrupp är också, tillsammans med gravt hörselnedsatta, i större utsträckning än övriga målgrupper vana att titta på direkttextade program.

De två filmklipp som fått högst snittbetyg har ett litet antal ord i texten. Testpersonerna uppgav att ett allt för stort antal ord i texten innebär att koncentrationen läggs på att läsa texten och att de då inte har möjlighet att se vad som händer i filmklippet. Detta blir extra påtagligt i prototypen då texten är placerad en bit under bilden vilket innebär att det är ännu svårare att läsa texten och samtidigt se vad som händer i klippet, framförallt när texten har ett stort antal ord. Flera testpersoner uppger att de därför måste välja om de koncentrerar sig på texten eller bilden.

*”Väldigt mycket ord, måste koncentrera sig på texten.”* – Testperson med grav hörselnedsättning

Testpersonerna reagerar på skillnader i hur texten presenteras i prototypen och Velotypeklippet. Flera påpekar under testet att det är enklare att läsa texten när den är rullande, till skillnad mot dagens teknik där texten försvinner och börjar om på första raden när de två raderna är fulla. Några reagerar på skillnader i de två direkttextningsmetoderna gällande ordens flöde i textningen. Text som produceras med Velotype-tangentbord visas på skärmen ord-för-ord, i den takt de skrivs. I prototypen presenteras orden ibland i sjok om flera ord, ibland ord-för-ord och ibland presenteras halva ord som sedan bildar hela ord. Flera testpersoner reagerade på att det är svårt att läsa texten när textens flöde och rytm ständigt varierar och att de därför föredrar hur texten presenteras med dagens Velotype-teknik.

---

*"Om det rullar eller inte spelar ingen roll så länge texten kommer i lagom takt."* –  
Testperson som inte har svenska som modersmål

Resultaten från användartesterna tyder även på att användarna anser att felen som uppstår med skuggtalartekniken är mindre problematiska än de fel som uppstår vid textning med ett Velotype-tangentbord. Vid användning av skuggtalarteknik blir felaktiga ord snarare än felstavade ord. Detta eftersom att inga felstavade ord finns inlagda i systemet. Med dagens teknik kan felen bli totalt oförståeliga, där det kan stå "sleon" istället för "solen". Men skuggtalartekniken skulle felet istället kunna vara "kjolen". Användarna uppger att det är enklare att förstå dessa fel. Intervjuer som Stelacon genomfört inom ramen för studien pekar också på att liknande observationer har gjorts i andra länder. Acceptanstester som genomförts på Lingsofts prototyp pekar på att träffsäkerheten för taligenkänningen är 98,23 procent. Detta innebär att direkttextningen kan komma att innehålla väldigt få fel och i det avseendet vara bättre än dagens teknik.

## 4. Effekterna för tv-bolag

Vilka effekter ett införande av skuggtalarteknik har för ett programföretag beror på många olika faktorer. Som beskrivits i omvärldsanalysen ovan så har olika länder olika sätt att arbeta med direkttextning. Vissa länder lägger endast ut korrekta texter, vilket har stor påverkan på fördröjningen. Andra har valt att ha en fördröjning på sändningen vilket innebär att det finns extra tid för textaren att texta inslaget. Då SVT i detta prototypstadium inte har en färdig plan för hur de skulle kunna organisera sig kring skuggtalarteknik, så är det svårt att säga exakt vilka effekter ett införande kommer att få. Stelacon erfar dock att en implementering framförallt kommer att få effekter för SVT:s interna organisation, och har liten inverkan på tittarnas upplevelse av direkttextningen.

En lärdom från andra länder är att en av de främsta positiva effekterna vid införande av skuggtalarteknik är att upplärning av nya direkttextare är betydligt kortare med den nya tekniken. Istället för en flerårig utbildning krävs nu endast ett par månaders upplärning innan en person kan arbeta som direkttextare. Detta innebär kostnadsbesparingar för programföretagen i ett antal hänseenden. Då det inte kräver lika tydliga specialkunskaper med en flerårig utbildning, kan antalet skuggtalare öka och kan leda till minskade lönekostnaderna för programföretag samt minskade rekryteringskostnader. Erfarenheter från andra länder indikerar även att det också troligtvis blir lättare att rekrytera nya personer vid personalbrist samt enklare att planera verksamheten och snabbt utbilda nya resurser vid exempelvis semesterar. Rent kostnadsmässigt finns det troligtvis även en potentiell besparing i de hänseende att utrustningen vid skuggtalarteknik kan vara billigare än andra jämförbara tekniker, detta bedöms dock endast ge marginella effekter.

Då skuggtalartekniken i Sverige fortfarande är i ett prototypstadium och både tekniken och skuggtalarna kommer att läras upp och bli bättre vid ytterligare användning av tekniken, så är det rimligt att anta att kvalitén på texten kommer att förbättras vid en eventuell implementering. Bättre kvalitet på texten innebär att programföretagen har möjlighet att, i större utsträckning än idag, reprisera direkttextade program utan att bearbeta textningen. SVT uppger att de idag bearbetar ungefär hälften av sina direkttextade program innan de sänds igen. Att slippa denna bearbetning skulle ytterligare innebära kostnadsbesparingar för SVT.

Vidare, om ett tv-bolag som SVT bestämmer sig för att texta även förinspelade program med skuggtalarteknik så kan produktiviteten öka, detta menar både Lingsoft och en expert för Ericson. Deras erfarenheter visar på en ökning i produktivitet med upp till 20-30 procent. Detta innebär att en minuts tal, som med den tidigare tekniken tagit cirka 10 minuter att texta, kan textas på sju till åtta minuter med skuggtalarteknik. Det vore också möjligt att fördela arbetet så att en mer senior kompetens ansvarar för slutbearbetning och juniorer arbetar med skuggtalningen.

Flera personer som Stelacon intervjuat inom ramen för studien har uppgett att skuggtalartekniken även innebär en bättre ergonomi för textarna. Istället för att skriva på ett tangentbord talar textaren in i en mikrofon, detta minskar risk för skador i rygg och nacke. Många hävdar att en blandning av de olika teknikerna skulle vara optimalt för att skapa en ergonomisk arbetssituation för textarna.

I samtal med olika branschaktörer och experter framkommer att om SVT skulle välja att gå vidare och fullskaligt implementera prototypen i sin nuvarande verksamhet skulle textarnas arbetsbeskrivning förändras och vissa personer kan eventuellt ha svårigheter att anpassa sig till de nya kraven.

---

Prototypen är framtagen för att texta vädersändningar och språkdatan är därför anpassad efter de vokabulär som används i sådana sändningar. Om SVT vill implementera skuggtalartekniken i andra typer av sändningar, exempelvis nyheter, krävs en utvärdering av vilken data som finns och vad som behöver adderas. Även om tekniken finns så behöver data anpassas efter det område tekniken ska användas inom. Att bygga upp en databas för ett område med betydligt bredare ordomfång är tidskrävande och innebär därmed en stor kostnad.

Skulle andra aktörer än SVT välja att implementera skuggtalarteknik i sina organisationer så innebär det en stor uppstartskostnad för företaget. Att köpa in systemet eller själva bekosta utvecklingen av systemet står i direkt relation med vilka effekter ett utbyte av tekniken har för aktören. Så länge det inte går att peka på några stora vinster med att införa skuggtalartekniken så antas få aktörer vara intresserade av att själva bekosta utvecklingen av tekniken i deras organisation. En anledning till att andra aktörer skulle välja att satsa på den nya tekniken är vid en eventuell ökning av kraven på andel textade program. Vid en sådan ökning skulle behovet av att snabbt kunna rekrytera personalstyrka bli större.

## 5. Slutsatser

Studiens resultat tyder på att tittarens upplevelse av direkttextning inte kommer att förändras avsevärt om SVT väljer att implementera skuggtalartekniken. Den skillnaden som tittaren kan märka vid byte till skuggtalarteknik är att texten presenteras i ett annat flöde och rytm än tidigare samt att felen i texten blir annorlunda. Acceptanstester av Lingsofts prototyp, som Stelacon tagit del av, visar även på att tekniken har hög träffsäkerhet och att fel i texten kan komma att minska vid en eventuell implementering av skuggtalarteknik.

Den tydligaste positiva effekten av ett införande av skuggtalartekniken är att utbildningen av textarna är betydligt kortare än vid användning av dagens teknik. Detta medför framförallt en ökad flexibilitet för tv-bolag som snabbt kan ta in nya resurser i organisationen vid resursbrist eller semestrar samt eventuellt minska lönekostnader. Omvärldsanalysen har även visat att skuggtalarteknik ger möjlighet att möta de ökande kraven på andel textade program.

Användartesterna visar på att vid direkttextning tycks textens fördröjning vara den aspekt som har störst påverkan på hur man upplever kvalitén på textningen. De filmklipp som textats med prototypen och som användarna fick se i användartesterna har en relativt kort fördröjning, de flesta på under fyra sekunder. Erfarenheterna från andra länder indikerar dock att skuggtalartekniken inte avhjälper detta problem, utan fördröjningen kommer troligtvis även fortsättningsvis vara snarlik de nivåer som fås vid användning av dagens Velotype-teknik. Generellt sett så visar erfarenheter från andra länder har tekniska lösningar svårt att minska fördröjningen. De enda tillförlitliga tillvägagångssätten är snarare:

1. **Förbereda material i större utsträckning.** Detta är framförallt möjligt vid nyhetssändningar där livesändningen blandas upp med förinspelat material. Om textarna i förväg får tillgång till det förinspelade materialet finns kan detta textas i förväg och endast delarna som sänds live direkttextas. Detta skulle innebära att den genomsnittliga fördröjningen för klippet minskar.
2. **Sätta en medveten fördröjning.** Genom att medvetet fördröja sändningen och ge textarna tillgång till sändningen direkt får textarna försprång och den fördröjning som tittarna upplever blir kortare. Vid användning av denna metod är det viktigt att kontrollera fördröjningen så att den inte blir för kort, det vill säga att texten kommer ut i sändning innan orden sagts. Detta kan framförallt bli problematiskt om vissa delar direkttextas och andra delar är förinspelade.

Givet behovet av minskad fördröjning så bör eventuellt SVT utvärdera vilken målgrupp deras textning vänder sig till. Om målgruppen är bredare än personer som har grav hörselnedsättning eller dövhet så bör ett övervägande göras om det finns möjlighet att göra någonting för att minska fördröjningen, exempelvis senarelägga livesändningar eller förbereda mer material för att minska fördröjningen. För personer som använder texten som stöd till ljudet så är det av extra stor vikt att ljud och text kommer i nära anslutning till varandra.

Vidare visar omvärldsanalysen att länder har valt att fokusera på olika aspekter i undertexternas kvalitet baserat på kulturella och språkliga förutsättningar. I Storbritannien återges ordagrant det som sägs, medan Belgien, baserat på ett relativt komplicerat språk valt att i stor utsträckning sammanfatta mycket av det som sägs. Samtidigt i Frankrike är det stort fokus på korrekta undertexter, vilket självklart får effekter på fördröjningen. Då skuggtalartekniken innebär förändrade förutsättningar i form av exempelvis andra typer av fel och möjlighet att snabbt producera stora mängder text finns det troligtvis ett behov på SVT att

---

genomlys dagens riktlinjer gällande avvägning mellan träffsäkerhet, fördröjning och rättningar i texten utifrån svenska behov och förutsättningar.