

MODIFIERAD STANDARDORTOGRAFI (MSO6)

Version 6
Januari 1999

Institutionen för lingvistik
Göteborgs universitet

Joakim Nivre¹

1 Allmänna principer

MSO6 är en modifierad standardortografi för svenska avsedd att användas vid transkription av talspråk. Den bygger på följande tre huvudprinciper:

- Ord som bara förekommer i en uttalsvariant transkriberas alltid med standardortografi. Således skrivs alltid "sverige" och "kanske", eftersom dessa ord normalt uttalas /'sverje/ respektive /'kanʃe/. (Uttalen /'sve:rige/ och /'kanske/ förekommer i praktiken inte.)²
- Ord som förekommer i flera uttalsvarianter transkriberas i enlighet med det aktuella uttalet genom modifiering av standardortografin. Således transkriberas ordet *jag* som "ja" eller "jag", beroende på om det uttalas med hörbart /g/ eller inte.
- I de fall då den modifierade standardortografin ger upphov till ambiguitet (homografi) som inte återfinns i skriftspråket, disambigueras homografa former med hjälp av klamrar eller numeriska index (se nedan). Således måste "ja" som en uttalsvariant av *jag* (pronomen) skiljas från "ja" som en uttalsvariant av *ja* (feedback-ord). Däremot görs inga försök att disambiguera former som är homografa även i standardortografi (t ex substantivet *vad* och pronomenet *vad*).

I avdelning 2 beskrivs kortfattat de olika uttalsvariationer som kan fångas med hjälp av modifierad standardortografi. I avdelning 3 beskrivs disambiguering med hjälp av klamrar ("krullparenteser"), vilket är den metod för som i första hand skall användas eftersom den underlättar såväl transkription som kontroll. I avdelning 4 beskrivs disambiguering med numeriska index, vilket måste tillgripas i de fall då disambiguering med klamrar ej är möjlig. (Listan över de ordformer som måste disambigueras på detta sätt återfinns i appendix.) I avdelning 5 ges slutligen en instruktion för hur man tillämpar MSO6 vid praktiskt transkriptionsarbete.

¹ Detta dokument är resultatet av ett kollektivt arbete med bidrag från många personer vid institutionen för lingvistik, Göteborgs universitet, däribland Jens Allwood, Elisabeth Ahlsén, Leif Grönqvist, Staffan Larsson, Sylvana Sofkova, Kristina Tullgren och Maria Björnberg.

² Skulle ett avvikande uttal trots allt förekomma, markeras detta med en uttalskommentar:
@ < pronunciation: 'sve:rige >.

2 Vad kan återges i modifierad standardortografi?

Användningen av modifierad standardortografi för transkription av talspråk gör det möjligt att fånga drag i talspråket som går förlorade vid transkription med standardortografi. Samtidigt är det viktigt att komma ihåg att det ingalunda är fråga om någon fonetisk, eller ens fonematisk, transkription, och att många uttalsnyanser helt enkelt inte går att fånga med detta transkriptionssätt. Därför vill i denna avdelning kort gå igenom vilka talspråksdrag som kan och skall markeras, och samtidigt varna för några fenomen som alltför ambitiösa transkriptörer ibland försöker återge trots att de ligger utanför systemets räckvidd och därför bara ger upphov till felaktigheter i transkriptionerna. Grovt sett kan vi dela upp de modifieringar av standardortografi som kan förekomma i tre grupper:

1. Reduktioner (dvs bortfall av grafem).
2. Substitutioner (dvs utbyten av grafem).
3. Övriga modifikationer.

2.1 Reduktioner

Den vanligaste typen av modifiering beror på att standardortografi ofta återger längre ordformer än de som finns i talspråk. Nedan följer ett antal exempel på transkriptionsformer med motsvarande standardform i högerspalten:

Transkriptionsformer

jag, ja, a
det, de, e
den, en
springer, springe, spring
kastade, kasta
roligt, roligt

Standardform

jag
det
den
springer
kastade
roligt

Fr o m MSO6 skall alla transkriptionsformer som kan fås från standardformen enbart genom att ta bort bokstäver transkriberas så att de utelämnade bokstäverna återges inom mängdklamrar ("krullparenteser"). Detta innebär att ovanstående exempel skall skrivas på följande sätt:

Transkriptionsord

jag, ja{g}, {j}a{g}
det, de{t}, {d}e{t}
den, {d}en
springer, springe{r}, spring{er}
kastade, kasta{de}
roligt, roli{g}t

Standardortografi

jag
det
den
springer
kastade
roligt

Observera att klamrar kan förekomma var som helst i orden och att klamrar kan innehålla flera bokstäver. I de fall då man inte säkert kan avgöra vilken standardform som är aktuell används *-former (se avdelning 3 nedan).

2.2 Substitutioner

En annan vanlig typ av uttalsvarianter kan beskrivas genom substitution, dvs genom att ett fonem/grafem byts ut mot ett annat. T ex är det vanligt i många dialekter att infinitivformens slut-/a/ byts ut mot /e/, vilket gör att standardformen *kasta* kan motsvaras av såväl "kasta" som "kaste". Andra exempel på substitutioner är "stog" (*stod*) och "ifjor" (*ifjol*).

I detta sammanhang kan det vara värt att påpeka att möjligheten att återge allofonisk variation (dvs variation mellan olika uttal av ett och samma fonem) är i det närmaste obefintligt inom modifierad standardortografi. Transkriptörer bör därför avhålla sig från att försöka återge t ex skillnader i öppenhetsgrad för vokaler eller olika realiseringar av sje-ljudet, eftersom det inte finns något entydigt ortografiskt sätt att koda de olika varianterna. Varianter som "bett"-"bätt" och "kanske"-"kansje" är således helt meningslösa och skapar enbart problem vid den efterföljande analysen av transkriptionerna. (Skriv alltid "bett" respektive "kanske".)

Låt oss också passa på att kommentera några högrekventa ord som ofta transkriberas fel i detta avseende:

- Konjunktionen 'och' och infinitivmärket 'att' uttalas normalt [o] i talspråk och skall då transkriberas "å" (mera specifikt "å0", eftersom formen "å" också kan svara mot standardformen *att*); formen "o" skall inte användas såvida inte det aktuella ordet uttalas med tydligt slutet kvalitet ([u:]), vilket torde vara mycket sällsynt.
- Verbformen *är* realiseras oftast utan /r/ och skall då transkriberas "e" om vokalen är halvsluten till halvöppen och/eller kort. Endast om vokalen är markant öppen ([æ:]) skall transkriptionsformen "ä{r}" användas.
- På samma sätt skall orden *den, det, dej, mej, nej, sej* alltid transkriberas med "e" (inte "ä") även i reducerade former, utom i de fall då vokalen är lång och markant öppen (troligen bara aktuellt i fallet "dä" för *det*).

2.3 Övrigt

Det stora flertalet av modifierade standardformer kan hänföras till någon av grupperna (enkla) reduktioner och substitutioner. Härutöver förekommer dels komplexa kombinationer av dessa operationer (t ex "kaste" för *kastat*, "re" för *det*), dels fall av inskott och tillägg (t ex "fiskarena" för *fiskarna* och "jaa" för *ja*). För dessa fall gäller att transkriptören bör försöka återge det aktuella uttalet så gott det går utifrån normala ortografiska principer (dvs att dubbeltecknad konsonant markerar kort vokal, att "k" betecknar /k/ före

bakre vokal osv). Dock gäller även här att man inte bör avvika från standardortografi mer än nödvändigt och inte återge allofonisk variation.

I detta sammanhang kan det också vara värt att påminna om förekomsten av s k oortodoxa standardformer, dvs former som inte alltid anses korrekta i skriftspråket men som måste räknas som standardformer i samband med talspråkstranskription. Således motsvaras pronomenet "dom" av standardformen *dom* (inte *de* eller *dem*), och formen "sa" ses inte som en reducerad variant av *sade*. Nedan följer en lista över de viktigaste fallen där tveksamhet kan uppstå om vilken norm som skall tillämpas. Listan skall läsas så att de former som är uppräknade i vänsterspalten kan användas i transkriptioner utan klamrar eller index (när uttalet så kräver); dessa former skall således inte ses som varianter av motsvarande form(er) i högerspalten.

Standardform:	Skall inte föras ihop med:
dej	dig
dom	de, dem
la	lade
mej	mig
nån	någon
nånting	någonting
nåt	något
nä	nej
sa	sade
sej	sig
ska	skall
sån	sådan
såna	sådana
sånt	sådant

3 Disambiguering med hjälp av klamrar

En grundläggande princip för MSO-transkribering är att alla modifikationer av standardortografi som ger upphov till ambiguitet (homografi) som inte återfinns i skriftspråket skall disambigueras. För reduktioner tillgodoses detta krav genom att *alla* reduktioner transkriberas så att de utelämnade grafemen tas med i transkriptionen men innesluts i mängdklamrar ("krullparenteser"). Följande exempel (från avdelning 2) illustrerar principen:

Transkriptionsord	Standardortografi
jag, ja{g}, {j}a{g}	jag
det, de{t}, {d}e{t}	det
den, {d}en	den
springer, springe{r}, spring{er}	springer
kastade, kasta{de}	kastade
roligt, roli{g}t	roligt

Fördelen med denna konvention är att transkriptören inte behöver hålla reda på vilka reduktioner som *de facto* ger upphov till nya ambiguiteter, eftersom den aktuella standardformen alltid kan rekonstrueras från en transkriptionsform med klamrar. Detta eliminerar också det behov av listor över ambigüosa former som uppstår vid disambiguering med hjälp av numeriska index (jfr avdelning 4). Det är dock viktigt att lägga märke till följande restriktion på användningen av klamrar:

- Klamrar kan inte användas då reduktion kombineras med substitution eller tillägg (jfr avdelning 2.3), eftersom man då inte entydigt kan rekonstruera standardformen från transkriptionsformen. Således måste "kaste" (för kastat) och "re" (för det) transkriberas utan klamrar.

Om behov av disambiguering uppstår i sådana fall, måste disambiguering ske med hjälp av numeriska index på det sätt som beskrivs i avdelning 4.

Om man är osäker på vilken den aktuella standardformen är (t ex därför att den avsedda betydelsen inte framgår i sammanhanget) används *-former på samma sätt som vid disambiguering med index (jfr avdelning 4). Om man således hör att en person säger "kasta" men inte kan avgöra om den avsedda formen är infinitiv, presens, preteritum, supinum eller perfekt particip skriver man "kasta*" (vilket således betyder att något *kan* vara utelämnat).

4 Disambiguering med hjälp av numeriska index

I de fall då en modifikation av standardortografi ger upphov till en ny ambiguitet (homografi) och den aktuella formen inte kan disambigueras med hjälp av klamrar, måste transkriptionsordet förses med ett numeriskt index som anger till vilken standardortografisk form ordet skall föras. Om exempelvis ordet 'och' transkriberas "å" (vilket motsvarar det vanligaste uttalet i talspråk), så uppstår ett sammanfall dels med förekomster av substantivet 'å', dels med förekomster av infinitivmärket 'att' vars vanligaste talspråksrealisering också bör transkriberas "å". Eftersom dessa tre ord skiljs åt i standardortografi, skall de också skiljas åt i transkriptioner, varför de tre transkriptionsorden skrivs "å0", "å4" respektive "å1". Observera att indexsiffran skall skrivas direkt efter ordet, utan mellanrum. Observera också att indexsiffran alltid skall stå sist i ordet, vilket innebär att eventuella förlängningstecken skall stå före siffran, alltså "å:0" (inte "å0:").

Som framgår av ovanstående exempel finns det ingen naturlig koppling mellan ett numeriskt index och motsvarande standardform (även om siffran "0" ofta används antingen för det mest frekventa ordet eller för det ord vars standardortografi sammanfaller med den indexerade talspråksformen). Av historiska skäl finns dessutom luckor i ordens indexserier som inte heller alltid börjar med "0". Följaktligen måste alla indexerade transkriptionsord förtecknas i en särskild lista, där det också

framgår till vilken standardform respektive index hör. Den aktuella listan för MSO6 ges som appendix till detta dokument. Observera att alla ord som återfinns i listan utan undantag måste indexeras vid transkription enligt MSO6.

I anslutning till listan kan det vara värt att uppmärksamma den speciella behandling som getts åt tveord som skrivs med ett enda grafem, närmare bestämt "e", "m", "u", "ä" och "ö". Dessa har genomgående begåvats med ett särskilt index, som skiljer dem från andra ord som skrivs på samma sätt (t ex substantivet 'ö'), trots att denna distinktion inte görs i standardortografi (i den mån tveorden över huvud taget förekommer i skriftspråk). Dessa ord har genomgående index 1 och skrivs följaktligen: "e1", "m1", "u1", "ä1" och "ö1".

Transkriptionsord som inte återfinns i listan, och som inte heller kan skrivas med klamrar (jfr avdelning 3), skrivs direkt i transkriptionen utan vare sig klamrar eller numeriska index. Dessa ord är då normalt antingen skrivna med standardortografi eller med en sådan modifikation av standardortografi att ingen ny ambiguitet (homografi) uppstår. Skulle man emellertid som transkriptör använda en ordform som ger upphov till en ny ambiguitet (och som varken kan disambigueras med klamrar eller befintliga index), skall man markera detta med hjälp av en särskild kommentar (se avdelning 5). Upptäckten av en sådan ambiguitet kan då leda till att listan över indexerade former utvidgas, men det är inte tillåtet för transkriptörer att på egen hand införa nya index (eftersom detta kan leda till sammanfall om flera transkriptörer inför nya index oberoende av varandra).

Om man inte kan avgöra vilket index som skall användas, därför att man inte kan avgöra vilken den avsedda standardformen är skrivs "*" i stället för index. Om man t ex hör att någon säger "kaste" men inte kan avgöra om den avsedda formen är presens, preteritum eller supinum av verbet *kasta*, så skriver man "kaste*".

5 Tillvägagångssätt vid transkription

Efter att ha beskrivit de allmänna principerna för transkription och disambiguering med MSO6, övergår vi nu till att ge en kortfattad vägledning för hur dessa principer tillämpas i praktiken. Vid transkription av ett inspelat samtal lyssnar man på bandet och skriver ner orden såsom de uttalas (givet MSO:s begränsningar; jfr avdelning 2) med iakttagande av följande regler:

- Om ordet kan skrivas med klamrar, skall det skrivas med klamrar. (Om man inte kan avgöra vilka de utelämnade grafemen är skrivs den reducerade formen följt av "*", t ex "kasta*").
- Om ordet inte kan skrivas med klamrar men förekommer i indexlistan gäller följande:

- Om den aktuella standardformen har ett index i listan, skall ordet skrivas med detta index (t ex "å0" för *och*).
- Om man inte kan avgöra vilken av de möjliga standardformerna som avses, skall ordet skrivas med en asterisk "*" i stället för indexsiffra (t ex "å*").
- Om den aktuella standardformen saknas i listan, skall ordet skrivas utan index men med kommentaren < SO: *standardform* >, där *standardform* är den aktuella standardformen. Detta gäller alltså då man upptäcker att en redan ambiguös transkriptionsform kan stå för ytterligare en standardform. (Meddela dessutom gärna ansvariga om upptäckten av en ny ambiguitet.)
- Om ordet inte kan skrivas med klamrar och inte förekommer i indexlistan gäller följande:
 - Om ordet uppenbart sammanfaller med ett ord från vilket det är skilt i skriftspråket, skall ordet skrivas utan index men med kommentaren < SO: *standardform* >, där *standardform* är den aktuella standardformen. Detta gäller alltså då man upptäcker en ny talspråksmodifiering som sammanfaller med en standardform. (Meddela dessutom gärna ansvariga om upptäckten av en ny ambiguitet.)
 - Annars skrivs ordet utan vare sig klamrar, index eller särskild kommentar.

Efter att en transkription har slutförts bör man extra noga kontrollera indexeringen av alla förekomster av ordtyper som har index, i synnerhet de vanligast förekommande (e, m, å). Detta görs enklast med användning av sök-och-ersätt-kommandot i det ordbehandlingsprogram som används.