Forskning om teckenspråk XXIV

Johanna Mesch

Svensk teckenspråkskorpus - dess tillkomst och uppbyggnad

Lars Wallin

Johanna Mesch

Annoteringskonventioner för teckenspråkstexter

2015

Institutionen för lingvistik

Stockholms universitet

Svensk teckenspråkskorpus – dess tillkomst och uppbyggnad 1

Johanna Mesch

Annoteringskonventioner för teckenspråkstexter 27

Lars Wallin och Johanna Mesch

Svensk teckenspråkskorpus – dess tillkomst och uppbyggnad

Johanna Mesch

Innehållsförteckning

1. Vad är en korpus? 1
2. Teckenspråkskorpusar 1
2.1. Vad är en teckenspråkskorpus? 2
2.2. ECHO-projektet 2
2.3. Internationellt nätverk för teckenspråkskorpusar 3
2.4. Teckenspråkskorpusar i andra länder 3
3. Korpus för det svenska teckenspråket 5
3.1. Förberedelser 5
3.2. Informanter 7
3.3. Material och inspelning 8
3.4. Redigeringsarbete 10
3.5. Annoteringsarbete 12
3.6. Etiska frågor 13
3.7. Metadata 15
4. Övriga datamängder 15
5. Tillgänglighet och användning 17
5.1. Teckenspråkslexikografi 18
5.2. Korpusbaserade studier i svenskt teckenspråk 20
5.3. Undervisningsmaterial 21
6. Slutord 23
Referenser 24

1. Vad är en korpus?[1]

Språkkorpusar finns för skrivna, talade och tecknade språk. En korpus är en sökbar textsamling, "*[a] collection of sampled texts, written or spoken, in machine-readable form which may be annotated with various forms of linguistic information*" (McEnery et al. 2006). För att en textsamling ska kunna kallas en korpus måste ett antal villkor uppfyllas:

- den innehåller naturligt språkbruk, autentiska texter som helst är representativa för en texttyp, genre, författare, tidsperiod e.d.
- den är stor, det vill säga den innehåller ett omfattande språkligt material
- den är datorläsbar, så att det går lätt att söka i och bearbeta materialet. [2]

Korpusar används för att empiriskt studera hur språk används. De kompletteras med introspektion där forskaren utgår från sin egen språkkänsla. Man måste komma ihåg att korpusar inte är heltäckande och att det finns mycket variation i hur språk uttrycks och används. Det är därför svårt att skapa ett representativt material som är tillräckligt omfattande eftersom språken är obegränsade då det gäller sina uttryckmöjligheter. Men en fördel med en korpus som består av begränsat och representativt datamaterial är att man kan studera *ett urval* av ett språk.

Korpusar är samplade, vilket innebär att någon form av medvetet urval har gjorts som samtidigt är representativt för en språk- eller texttyp. Materialet är också autentiskt vilket dock inte innebär att det måste vara i sin ”ursprungliga” form, utan om de till exempel är talade eller tecknade så måste de transkriberas. Då det gäller det svenska språket består enligt Borin (2011) den nationella korpusen av 90 % skriftspråk och 10 % talspråk.[3] Internationellt är den mest kända ordklasstaggade korpusen Brown corpus (1961) som är baserad på det engelska skrivna språket.

Att bygga upp en korpus kräver en hel del resurser, noggrann planering och arbete i olika faser. I denna uppsats kommer uppbyggnaden av en korpus för det svenska teckenspråket att fokuseras och beskrivas.

2. Teckenspråkskorpusar

I detta avsnitt kommer begreppet teckenspråkskorpus att tas upp och diskuteras. Under 2000-talet har den korpuslingvistiska inriktningen inom teckenspråksforskningen ökat kraftigt. Det så kallade ECHO-projektet (se avsnitt nedan) gav under 2003-2004 en tydlig riktning för hur en

---

[1] *Tack till Lars Wallin, Ulla Bjursäter och Ingela Holmström för synpunkter och språkgranskning av artikeln.*

[2] *http://sprakteknologi.se/vad-aer-sprakteknologi/lexikon/korpusar*

[3] *Det finns flera korpusar för det svenska språket, varav många finns samlade hos Språkbanken, Göteborgs universitet, se http://spraakbanken.gu.se/swe/resurser.*

teckenspråkskorpus ska se ut och hur den kan användas. SLCN-nätverket (se avsnitt nedan) under 2009-2010 blev sedan en inspirationskälla för den fortsatta utvecklingen av teckenspråkskorpusar. Flera länder har idag pågående, avslutade eller planerade projekt om teckenspråkskorpusar, däribland Sverige.

2.1. Vad är en teckenspråkskorpus?

Teckenspråkskorpusar liknar korpusar för talade språk såtillvida att manuella inmatningar av uttryck måste utföras. Medan en talkorpus innehåller ljud tillsammans med en transkription av vad som sägs, så innehåller en teckenspråkskorpus en videoinspelning tillsammans med en transkription av vad som tecknas. Alla tecken och övriga uttryck transkriberas (eller annoteras[4]) till kategorierna innehåll och funktion. För att en samling av teckenspråkstexter ska kunna kallas en teckenspråkskorpus måste följande villkor uppfyllas:

- texten ska vara datorläsbar (med hjälp av glosor och översättning),
- enheter ska vara sökbara med hjälp av dator (t.ex. med annoteringsverktyget ELAN)
- språket ska vara sådant som är naturligt förekommande och representativt (avseende faktorer som region, ålder, kön och skolgång)
- materialet ska bestå av olika texttyper, t.ex. fria samtal och eliciterat material
- materialet ska vara tillgängligt genom någon server eller webbportal

Teckenspråkskorpusar kan användas för att studera olika aspekter av tecknade språk, såsom pragmatiska, syntaktiska, semantiska och fonologiska. Man ska dock vara försiktig med att generalisera utifrån korpusmaterial, men de kan användas för att visa olika fenomen, hur olika tecken används i särskilda situationer och i varierande sammanhang. Korpusar kan också användas för att söka tecken i stora textsamlingar.

2.2. ECHO-projektet

Pionjärprojektet för annotering av teckenspråkskorpusar var ett EU-finansierat kulturarvsprojekt (European Cultural Heritage Online), som under projektnamnet ECHO-projekt leddes av Onno Crasborn under åren under 2003-2004. Forskare från Nederländerna, Storbritannien och Sverige samlade in teckenspråksmaterial som annoterades och publicerades på Internet. ECHO-projektet var det första i sitt slag i världen och det visade bland annat hur modern digital teknik erbjuder möjligheter att annotera och utbyta videomaterial mellan forskare som har tillgång till databasen (Crasborn 2004, Crasborn et al. 2004, 2007). ECHO-projektets material består av fem fabelberättelser, teckenspråkspoesi, samt korta intervjuer och översättningar till teckenspråk av en lista på 300 ord från tre teckenspråk: svenskt, nederländskt och brittiskt.

Under ECHO-projektet testades annoteringsverktyget ELAN för första gången för tre teckenspråk. Detta annoteringsverktyg visade sig erbjuda möjligheter att via dator transkribera teckenspråkstexter så att de kunde lagras i skriftformat i databasen. Verktyget är konstruerat så att de transkriberade texterna är länkade till videosekvenser. När ett transkriberat tecken eller en

[4] *Att annotera innebär att man skriver ned, antecknar det som sägs på olika nivåer.*

transkriberad teckenspråkssekvens markeras, kommer man direkt till motsvarande plats i den länkade videofilen. Teckenspråkstexter kan därmed lagras både i videoformat och i skriftformat, vilket har bidragit till förverkligandet av en teckenspråkskorpus. Inom ECHO-projektet skapades även det första förslaget till en standard för metadatabeskrivningar för teckenspråksdata. Via metasökningen kan den sökande navigera genom metadata och snabbt hitta en informationskälla. Svenskt material från ECHO-projektet är tillgängligt via språkarkivet vid Max Planck Institute ECHO - Sign Language[5] och genom Insitutionen för lingvistik, Stockholms universitet.[6]

2.3. Internationellt nätverk för teckenspråkskorpusar

Det europeiska nätverket *Sign Linguistics Corpora Network*, SLCN, verkade under åren 2009-2010 med finansiering från the *Netherlands Organisation for Scientific Research*, NWO. SLCN var ett nätverk som hade till syfte att samla befintliga kunskaper och metoder kring skapande och användning av teckenspråksresurser, att sprida denna information genom seminarier, att bygga en stabil webbplats för sådan kunskap och att skapa nya idéer kring användandet av den senaste tekniken för att studera tecknade språk. SLCN arrangerade fyra workshops kring datainsamling, metadata, annoteringar och användning.

Samarbetet i SLCN-nätverket resulterade bland annat i en webbportal för teckenspråkskorpusar. Webbportalen innehåller arbetsflödesbeskrivningar, etiska förfaranden och överväganden samt språkliga strategier för korpussamlingar, såsom annoteringskonventioner.[7] Detta ledde också till flera nya samarbeten, utveckling av standardannotering av teckenspråksdiskurser, samt utveckling av specifikationer för programvaror som passar för teckenspråkskorpusar och deras infrastruktur.

Efter nätverksprojektet har samarbetet fortsatt i olika former och vid olika workshops, bland annat *Digging into Signs workshop* som hölls i London i slutet av mars 2015. Där presenterade forskningsgrupperna annoteringskonventioner i sina korpusar för arton teckenspråk och diskuterade standard för framtida annoteringskonventioner.[8]

2.4. Teckenspråkskorpusar i andra länder

Det finns ett stort globalt intresse för teckenspråkskorpusarbetet. Annoteringsarbetet och sökmöjligheterna i multimodala annoteringsverktyget ELAN har lett till nya inriktningar för teckenspråksforskningen på ett internationellt plan. I många länder är teckenspråkskorpusar till stor del fortfarande under utveckling, de flesta är inte tillgängliga via nätet, utan ger bara information om de egna korpusprojekten.

Några av de främsta teckenspråkskorpusarna som finns publicerade är:

---

[5] https://corpus1.mpi.nl/ds/asv/;jsessionid=05D0ED27587252E33005B6970520E43F?0&openpath=node:84302

[6] http://www.ling.su.se/teckenspr%C3%A5ksresurser

[7] http://www.signlanguagecorpora.org

[8] http://www.bslcorpusproject.org/events/digging-workshop/

- ECHO-projektets korpusar för brittiskt teckenspråk (BSL), svenskt teckenspråk och nederländskt teckenspråk (NGT), 2004[9]
- Corpus NGT för nederländskt teckenspråk, 2008[10]
- BSL Corpus Project för brittiskt teckenspråk, 2011[11]
- The Auslan Corpus för australienskt teckenspråk, (avslutat 2008, tillgänglig 2012)[12]
- Svensk teckenspråkskorpus (avslutat 2011, tillgänglig 2012)
- DGS-korpus för tyskt teckenspråk, från 2009[13]

Pågående korpusarbete finns bland annat i Polen, Italien, Irland, Frankrike, Belgien, USA, Japan, Danmark, Finland, Norge, Slovenien och Spanien. Det har också gjorts omfattande insatser för att utveckla former av automatisering av teckenspråksmaterial, teckenspråksavläsning, rörelsespårningssystem och animerade avatarer (se t.ex. presentationer vid återkommande workshops under LREC-konferenser[14]).

De olika teckenspråkskorpusarna skiljer sig från varandra när det gäller språkvetenskapliga utgångspunkter, antalet informanter, antalet videokameror, elicitationsmetoder och inspelningsplatser under korpusbyggnandet. I de teckenspråkskorpusar som nämnts ovan har ett stort antal informanter i samtal annoterats, mestadels i form av glosor. Antalet informanter varierar mellan 40 och 300. Den drivande kraften och föregångaren inom teckenspråkskorpora har varit Nederländerna med sin korpus *Corpus NGT* (2006-2008).[15]

Den brittiska korpusen *BSL corpus* har en mer sociolingvistisk inriktning, medan den tyska korpusen *DGS-korpus* har ett stort och långsiktigt projekt med stark bas i iLex[16] som är ett verktyg för teckenspråkslexikografi och korpusanalys. Den irländska teckenspråkskorpusen *Ireland Digital Corpus in Irish Sign Language* i sin tur används främst vid undervisning i teckenspråkstolkning och för sociolingvistiska studier, medan den australienska *Auslan Corpus* är ledande när det gäller korpusbaserat teckenspråkslexikon, annoteringssystem och korpuslingvistiska sökfunktioner. Den australienska korpusen har också med mer syntaktisk information i sina annoteringar.

---

[9] http://www.let.ru.nl/sign%2Dlang/echo/

[10] http://www.ru.nl/corpusngtuk/introduction/welcome/

[11] http://www.bslcorpusproject.org/cava/

[12] http://www.auslan.org.au/about/corpus/

[13] http://www.sign-lang.uni-hamburg.de/dgs-korpus/

[14] http://www.sign-lang.uni-hamburg.de/lrec2012/workshops.html

[15] http://www.mpi.nl/resources/tools

[16] http://www.sign-lang.uni-hamburg.de/dgs-korpus/index.php/welcome.html
se också http://www.sign-lang.uni-hamburg.de/ilex/

3. Korpus för det svenska teckenspråket

Behovet av en korpus för det svenska teckenspråket lyftes bland annat fram i SOU 2006:29, som handlar om teckenspråkets ställning i Sverige:

> För svenska språket, i såväl skriven som talad form, finns numera allmänt tillgängliga databaser med stora textmängder. Sådana korpora utgör i dag oumbärliga verktyg för många olika typer av studier av svenska språket. Något motsvarande ”arkiv” med löpande text på teckenspråk finns inte och det pågår ingen dokumentation av olika typer av teckenspråksanvändning. Detta uppmärksammas av utredningen ”Mål i mun” som föreslår att resurser skapas för ”korpusarbete med teckenspråket” (SOU, 2002:27, s. 507). Ett brett upplagt korpusarbete behövs för såväl lexikonarbetet, som för fördjupad forskning om teckenspråkets struktur, men också för barnspråksforskning och forskning om tolkning mellan teckenspråk och svenska. (Ahlgren & Bergman, 2006:29, s. 57)

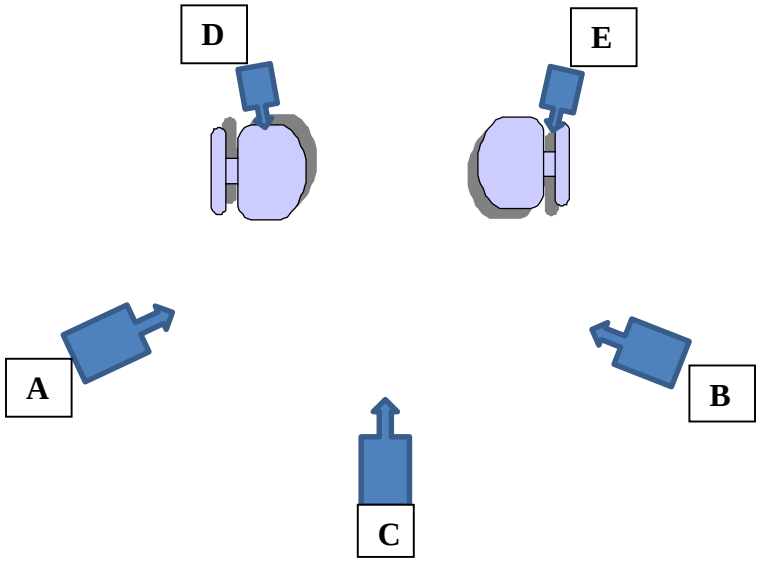
Mot denna bakgrund finansierade Riksbankens jubileumsfond åren 2009-2011 ett projekt med namnet *Korpus för det svenska teckenspråket*[17] vid Avdelningen för teckenspråk, Institutionen för lingvistik på Stockholms universitet.

Mellan juni 2009 och mars 2011 genomfördes inspelningar med 42 informanter i åldrarna 20 till 82 år från olika regioner i Sverige. Påföljande fas innebar ett intensivt redigerings- och annoteringsarbete av videomaterialets ca 167 timmar (inkl. olika bildvinklar). Annoteringsarbetet pågår fortfarande, där varje informants inspelning annoteras med glosor (tecken) och översättningar. Korpusarbetet beskrivs kortfattat i de kommande avsnitten: a) förberedelser, b) informanter, c) material och inspelning, d) redigeringsarbete, och e) annotering. Dessutom berörs etiska frågor och metadata.

3.1. Förberedelser

När projektet *Korpus för det svenska teckenspråket* startade den 1 januari 2009, påbörjades ett arbete med att förbereda för datainsamlingen. Inspelningsstudion rustades upp, riktlinjer utarbetades för annoteringar (dvs. vilka rader och glosor som skulle användas för korpusmaterialet), informanter kontaktades och några testinspelningar gjordes för att se om upplägget fungerade som tänkt och för att fastställa om rekvisita, kamerautrustning och belysning var tillräckligt bra. Testinspelningarna användes också för att se om eliciteringsmaterialet fungerade och för att komma fram till vilka bildvinklar som skulle användas. Samtliga inspelningar planerades till inspelningsstudion vid Institutionen för lingvistik, med fem kameror som var riktade på olika sätt (Fig.1-2).

[17] *Projektet ”Korpus för det svenska teckenspråket”, Riksbankens Jubileumsfond (In2008-0276-1-IK)*



Figur 1. Fem kameror och två informanter ur fågelperspektiv

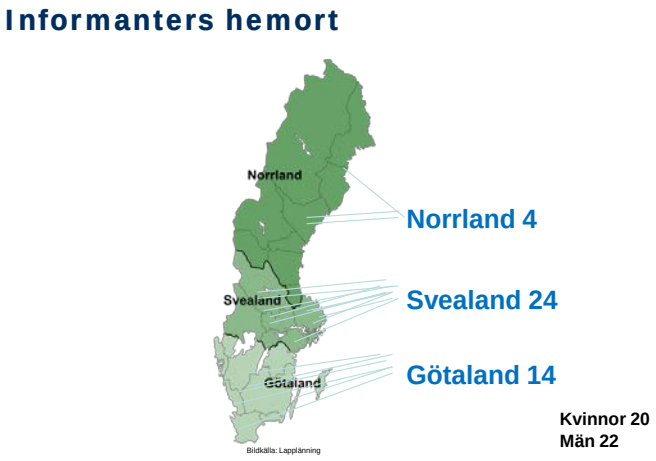
I förberedelsearbetet formulerades också en arbetsplan för korpusprojektet i sin helhet. Denna innehöll en tidsplan, en beskrivning av studio och kamerautrustning, ett upplägg för testinspelningar, en plan för kontakter med informanter, riktlinjer för utarbetande av annoteringskonventioner, etc. Dessutom gjordes en plan för hur korpusmaterialet skulle läggas upp i databasen och hur det skulle publiceras på Internet.

Figur 2. Inspelningsstudio

3.2. Informanter

För att få en korpus bestående av ett representativt material gjordes ett urval av informanter från olika åldersgrupper och regioner. Ett av de främsta kriterierna var att de skulle vara fullt teckenspråkiga och använda svenskt teckenspråk som sitt primära språk. Varje informant fick ta med sig en samtalspartner som kunde vara t.ex. en arbets-, skol- eller idrottskamrat. Innan inspelningarna påbörjades gicks etiska frågor igenom, som bland annat behandlade problematiken kring att det inte är möjligt att vara anonym vid publicering av korpusen på Internet. Informanterna blev också informerade om projektet på både svenskt teckenspråk och svenska, och de fick skriftligen ge sitt samtycke innan inspelningen startades. För att samtalen skulle flyta mer naturligt samtalpartnerna emellan inleddes inspelningarna med småprat. I studion deltog två döva medarbetare för att se till att inspelningarna skulle flyta problemfritt.

Alla informanter som har deltagit i projektet är födda i Sverige utom en, som flyttat till Sverige i mycket tidig ålder. Sex av informanterna hade minst en utländsk förälder och alla har olika familje- och språkbakgrund. Tolv av deltagarna hade döva föräldrar och tolv hade döva eller hörselskadade syskon. Informanterna hade varierande grad av hörsel, två var hörande och barn till döva föräldrar. Vidare hade informanterna gått i skolan under olika perioder, med eller utan teckenspråk i undervisningen. De hade också varierande utbildningsbakgrund och yrkesutbildning, med allt från folkskola/dövskola till universitetsutbildning. Olika yrkeskategorier var representerade, såsom pensionärer, studenter, montörer och lärare. De flesta var aktiva i döv- och teckenspråksrelaterade frågor i samhället. Den regionala fördelningen av de 42 informanterna illustreras i Fig. 3 och deras skolbakgrund i Fig. 4. Informanterna är någorlunda jämnt fördelade i sex åldergrupper, vilka illustreras i tabell 1.



Figur 3. Regional fördelning av informanterr

Figur 4. Informanternas skolbakgrund

Tabell 1. Informanternas ålder vid inspelningstillfället

	20-29	30-39	40-49	50-59	60-69	70-
Totalt	9	6	7	8	8	4

3.3. Material och inspelning

Hösten 2009 påbörjades själva korpusinspelningarna och fram till mars 2011 genomfördes 21 inspelningar med informanter. Eftersom varje inspelning har två informanter som samtalar med varandra blev det totala antalet inspelade informanter 42 personer. Varje inspelningstillfälle tog ungefär två timmar och omfattade bland annat en kort presentation, semi-spontana eller helt spontana fria samtal, upplevelseberättelser samt eliciterade berättelser. Alla samtal skedde som parsamtal.

Inspelningarna var upplagda så att inspelningsledaren inledde med en kort presentation av projektet på teckenspråk. Sedan ombads informanterna att presentera sig med namn, persontecken, nuvarande yrke och lite kort om sin uppväxt och skolgång. Efter detta ställdes några inledande frågor inom ett ämne eller flera delämnen som informanterna sedan själva kunde vidareutveckla. Några par började direkt samtala medan några behövde lite hjälp att komma igång, i form av fler följdfrågor eller nya frågor för att utveckla sina samtalsämnen. Frågorna var förberedda, men anpassade efter samtalsämnena under inspelningen, såsom ”Kan dina föräldrar teckenspråk?”, ”När börjar du teckna?”, ”Är/har du varit aktiv inom idrott eller kultur?”, ”Hur ser kommunikationen ut på din arbetsplats?” etc. Samtalen handlade oftast om dövrelaterade ämnen men också om andra samhällsrelaterade ämnen. Efter de här samtalen fick informanterna informanterna berätta om en viktig händelse i livet eller om något roligt minne. Samtalslängden varierade mellan 1 och 16 minuter per diskussionsämne och den genomsnittliga längden för samtalen var 40 minuter.

Efter en paus fick informanterna fortsätta med eliciterade berättelser och samtal. De fick titta i en bilderbok utan text, som *Frog, where are you*[18]*?* eller *The Snowman*[19] i 15 minuter. Efteråt fick de återberätta innehållet för varandra. De skulle återberätta innehållet så utförligt de kunde men behövde inte komma ihåg alla detaljer. Nästa uppgift var att återberätta bildserien *Fågelholken* (Informant A) och *Hönshuset* (Informant B) samt två valfria serier som Ferninand (se figur 5-8). Informanterna fick då titta på serierna i några minuter och sedan återberätta dem för varandra. Därefter fick informanterna titta på ett filmavsnitt om ca fem minuter, *Plankan* eller *Mr Bean* (2 av 42 använde *Pingu)*, och återberätta innehållet i filmavsnitten för varandra. Slutligen skedde fria samtal om filmer, TV-program eller andra diskussionsämnen.

Figur 5. Bilderböcker "Frog, where are you?" (Var är du, grodan?) och "The Snowman" (Snögubben)

Figur 6. Fågelholken och Hönshuset från NGT-korpusprojektet

[18] *Mercer Mayer (1969), Frog, where are you?*

[19] *Raymond Briggs (1978), The snowman*

Figur 7. En av de valfria serierna, Ferdinand-serie

Figur 8. Några korta filmsekvenser från Plankan och Mr Bean

3.4. Redigeringsarbete

Efter inspelningarna tog redigeringsarbetet av det inspelade materialet vid. Eftersom kamerorna som användes hade olika inspelningsfomat (hårddisk, minneskort och HDV-band) behövdes en del efterbehandling, därefter överfördes allt till redigerbar form i FinalCut Pro (och även Adobe Premiere). Det inspelade materialet redigerades och komprimerades till datafiler i form av MPEG-1, som är ett standardformat för ELAN och som fungerar på både PC och Mac-datorer. Totalt genererade inspelningarna 24 timmar färdigredigerade filmer för annotering. Videofilerna arkiverades i originalformat i MPEG-1 och senare även som MPEG-4. Varje videofil fick en egen filbeteckning med följande information:

SSLC = Swedish Sign Language Corpus
01 = inspelningsfasen som innehåller presentation och samtalsmaterial
02 = inspelningsfasen som innehåller eliciterat material
_001 = inspelningstillfälle (session)
_S001 = informants identitetskod
_b = bildvinkel för halvbild

En videosekvensfil kan alltså exempelvis betecknas SSLC01_001_S001_b. De olika bildvinklar som användes var halvbild, närbild, helbild och fågelperspektiv. Halvbild är ett bildutsnitt där en informant visas från höften och uppåt. Närbild är ett bildutsnitt på ansiktet och helbild är ett bildutsnitt där två informanter syns i sin helhet vända mot varandra. Det finns dessutom en

bildvinkel i fågelperspektiv genom kameraupptagning i taket. Den sista bokstaven efter informanten anger vilken bildvinkel som används, b=body, f=face, tu=top up, p=par.

Figur 9. S001_b (halvbild på S001)

Figur 10. S002_b (halvbild på S002)

Figur 11. S001_S002_p (helbild på S001 och S002)

Figur 12. S001_f (närbild på S001)

Figur 13. S002_f (närbild på S002)

Figur 14. S001_tu (fågelperspektiv på S001)

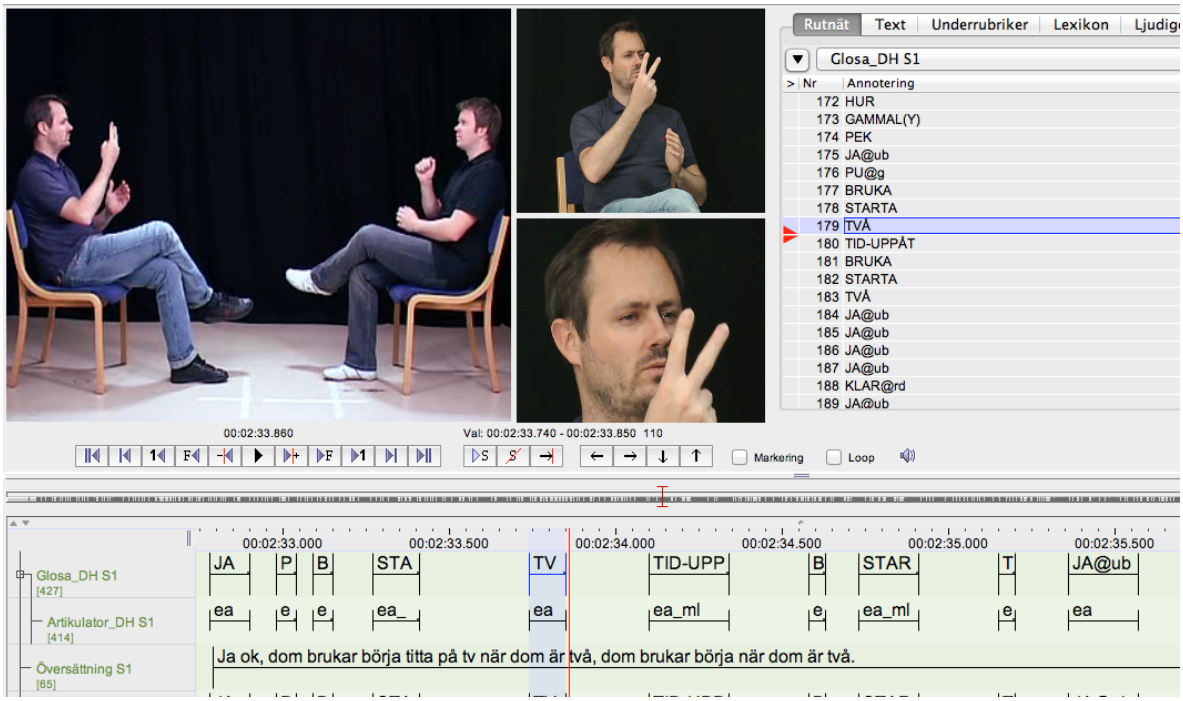
Figur 15. S002_tu (fågelperspektiv på S002)

3.5. Annoteringsarbete

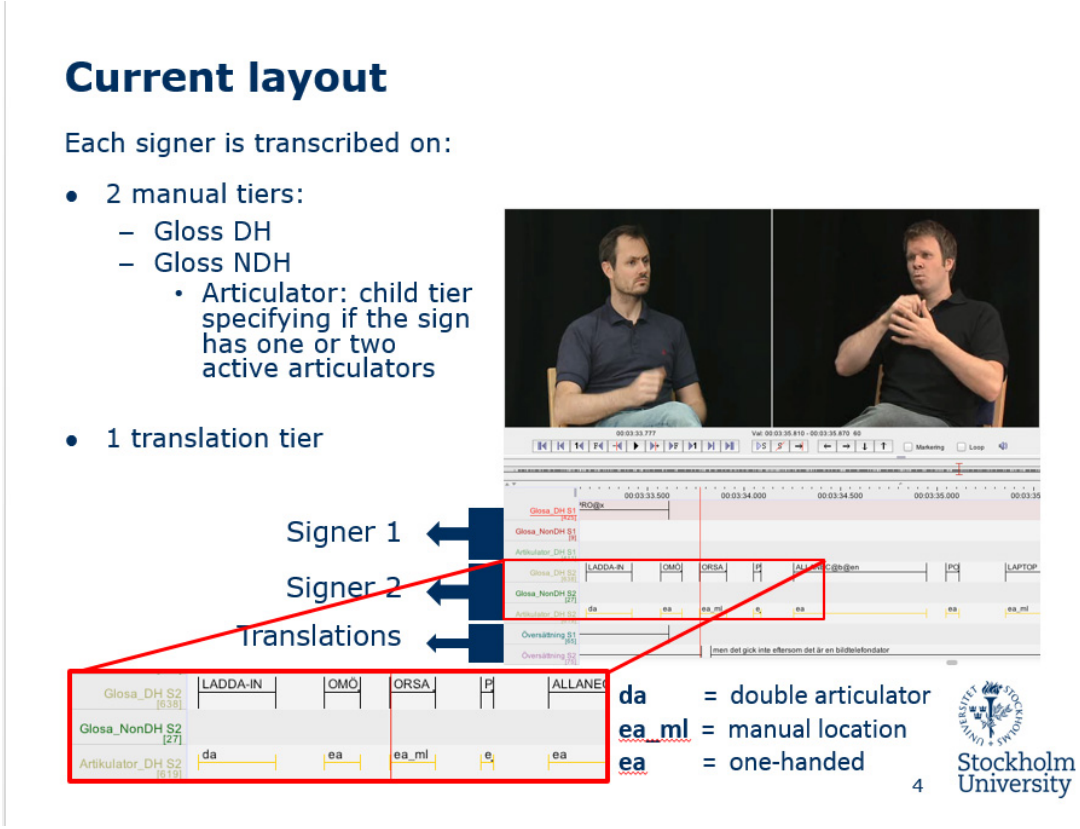
En teckenspråkskorpus är, som framgått hittills, uppbyggd av videofiler och annoteringar. Det behövs därför också annoteringskonventioner för teckenspråkstexter som underlättar annoteringsarbetet och sökning av glosor. Sådana konventioner har utarbetats inom korpusprojektet och beskrivs närmare i Wallin & Mesch (2014). Den senaste versionen är med i denna FOT-rapport. Utöver detta har bruksanvisningar för annoteringsverktyget ELAN publicerats (se Mesch 2011). För en utförlig presentation av annoteringskonventionerna hänvisas till dessa beskrivningar. (Se även Mesch & Wallin, 2015).

Annoteringsarbetet har utförts med hjälp av annoteringsverktyget ELAN. Exempel på hur ett annoteringsfönster i ELAN kan se ut visas i figur 16. Att annotera texter är, som tidigare nämnts, tidskrävande. Det tar ca 1 ½ timme att annotera glosor för en videosekvens på en minut. Sedan tar det ytterligare ca 1 timme att göra en svensk översättning på samma videosekvens (Nilsson & Rohdell, 2011). Annoteringsrader är ordnade efter olika radtyper, bland annat för glosor och funktion hos den icke-dominanta handen, se fig. 17.

Ett sätt att underlätta annoteringsarbetet är att skapa ett kontrollerat ordförråd (*Controlled Vocabulary, CV*), vilket innebär att annoteringstiden för en videosekvens på 1 minut för glosor kan reduceras till ca 1 timme. Annoteringsfilerna kallas för eaf-filer och de innehåller språkliga typer, rader och annoteringar. Eaf-filen läses tillsammans med den eller de videofiler den är länkad till. I korpusprojektet har ett informantpar 5-7 eaf-filer vardera.



Figur 16. Ett ELAN-fönster med rader för glosor och översättning



Figur 17. Överordnad rad för glosor och underordnad rad för funktion hos den icke-dominanta handen. En presentationsbild från föreläsningen på workshop i Reykjavik. (Börstell, Mesch & Wallin, 2014)

3.6. Etiska frågor

Institutionen för lingvistik på Stockholms universitet har upphovsrätten till det korpusmaterial som beskrivits här, men tillåter att korpusanvändare fritt får kopiera och annotera materialet vidare. Den som gör det måste dock följa de villkor som följer genom den licens för upphovsrätt, *Creative Commons*, som har använts (Fig. 18). Korpusanvändare får heller inte inte använda materialet i kommersiella sammanhang.

Figur 18. Vinjett till början och slutet av filmen avseende licensen för Creative Commons.

Samtliga informanter har undertecknat ett avtal för användning av inspelningarna till korpusprojektet. Att handskas med etiska aspekter av teckenspråkiga data är helt annorlunda än för korpusar av skrivna och talade språk, eftersom teckenspråket är visuellt och det krävs videoinspelningar av hela personen som tecknar. Det är svårt att dölja informanternas ansikten och därför kan teckenspråkiga informanter anses vara mindre anonyma än informanter i skrift- eller talspråk (se även Crasborn, 2010).

3.7. Metadata

Varje annoterad videofil i korpusen innehåller också metadata bestående av information om inspelningsdatum, mediatyp, texttyp, eliciteringsmetod, tillstånd, informant-ID, annoteringstyp, annotatör, projektmedverkande, syfte, tekniska metadata etc. Dessa uppgifter följer *the Component MetaData Infrastructure* (CMDI)[20].

Metadata används för att skapa bättre tillgänglighet och samarbete mellan olika universitet avseende datamaterial. Det finns färdiga metadatakomponenter i CMDI och även anpassade metadatakriterier för teckenspråksmaterial. Tack vare metadata kan speciell information sökas upp vilket medför en bättre överblick över videomaterialet, med bättre ordning och tillgänglighet. En webbportal med två skilda tillträden har skapats, ett slutet som är till för forskare och ett öppet som är till för allmänheten, såsom språkinlärare och språkutbildare.

Korpusmaterialet för det svenska teckenspråket har kategoriserats enligt datatyper och regioner.

Datatyper:	Presentation
	Samtal
	Eliciterade berättelser
Regioner:	Götaland
	Norrland
	Svealand

Ytterligare metadatabeskrivningar som avser ålder, kön, skola, språk, dövhet, ålder för inlärning av teckenspråk, högerhänthet etc. finns tillgängligt enbart för forskare.

4. Övriga datamängder

På lång sikt ska olika slags teckenspråkskorpusar och datamängder samlas i databasen. ECHO-filerna finns i svensk version enligt svenska annoteringskonventioner. Valda video- och annoteringsfiler av korpus för taktilt teckenspråk kommer att finnas tillgängliga på webben, Fig. 19. En del videodata som är mer forskningsriktade består av barns teckenspråk, Fig. 20, och korpus i svenskt teckenspråk som andraspråk, Fig. 21. Dessutom återfinns Anna-Lena Nilssons datamaterial på ca 13 timmar för VR-projekt i teckenspråkstolkning *Use of signing space in simultaneous sign language interpretation*. Dessa är s.k. slutna korpusar som är tillgängliga för forskare med särskilt tillstånd, se Fig. 22. Andra typer av material kommer att samlas in till korpusfamiljen, bland annat teckenspråkspoesi, föredrag och narrativa texter.

[20] http://www.clarin.eu/content/component-metadata

Figur 19. Datamängd för taktilt teckenspråk på ca 4 timmar (Mesch, under arbete).

Figur 20. Datamängd för barns teckenspråk på ca 100 timmar (Simper-Allen, under arbete).

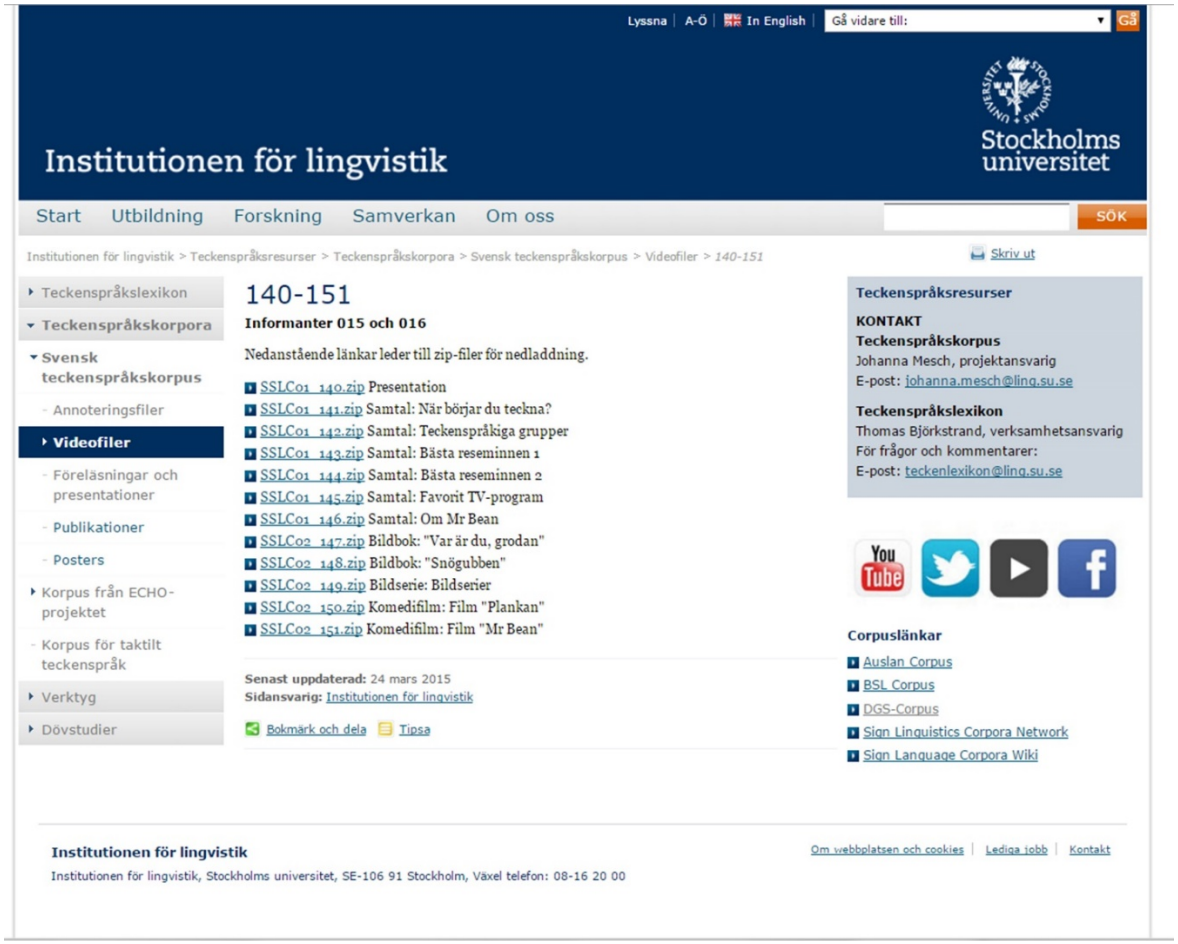
Figur 21. Datamängd för svenskt teckenspråk som andraspråk på ca 9 timmar (Mesch & Schönström, under arbete)

Figur 22. Teckenspråkskorpusar vid Institutionen för lingvistik, Stockholms universitet.

5. Tillgänglighet och användning

Den svenska teckenspråkskorpusen består av en avsevärt lägre andel text och förekomster än i skrivna korpora, men det material som hittills har skapats ger ändå en bra bild av hur teckenspråksmeningar ser ut och bidrar med nya tecken och varianter till Svenskt teckenspråkslexikon. Korpusen kan användas för att utveckla undervisningsmaterial i teckenspråk erbjuder också en möjlighet att presentera eller analysera enskilda tecken, meningar eller hela texter vid språkundervisning.

Korpusen är tillgänglig för forskare och undervisande personal vid Avdelningen för teckenspråk, Institutionen för lingvistik vid Stockholms universitet. Korpusmaterialet har också samlats till en hemsida som består av annoterade filer, vilka är fritt tillgängliga för allmänheten och den kan användas t.ex. i teckenspråks- och tolkutbildningar i Sverige. Nästan alla videofilerna är tillgängliga på hemsidan http://www.ling.su.se/teckensprakskorpus (Fig. 23). Där uppdateras annoteringsfiler (eaf-filer) med glosor och svensk översättning då och då på webben. En del av korpusmaterialet arkiveras i språkarkivet vid Max Planck Institute for Psycholinguistics, Nijmegen, Fig. 24.



Figur 23. Korpusdata på hemsidan för Institutionen för lingvistik, Stockholms universitet

Figur 24. Språkarkivet vid Max Planck Institute for Psycholinguistics, Nijmegen

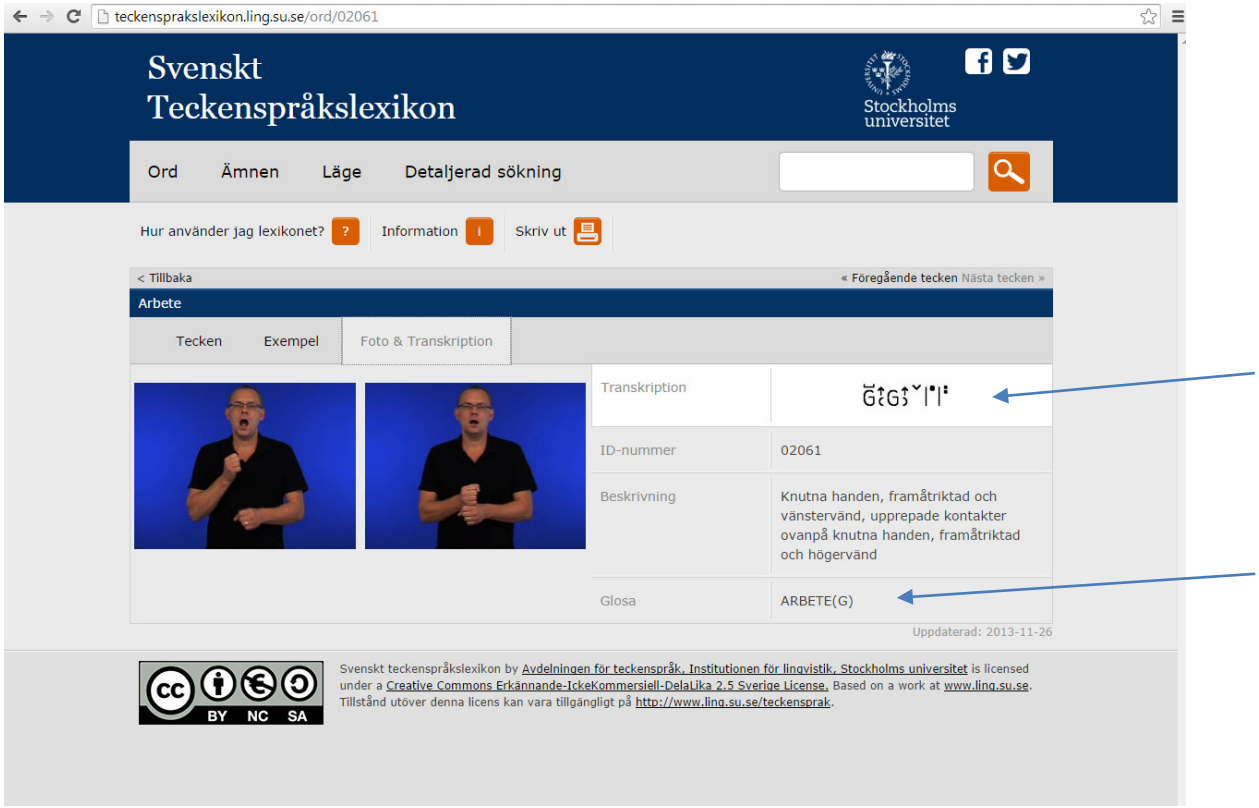
5.1. Teckenspråkslexikografi

Teckenspråkslexikografi är ett exempel på hur korpusundersökningar kan användas. Den lexikografiska verksamheten vid Avdelningen för teckenspråk på Institutionen för lingvistik vid Stockholms universitet har som huvuduppdrag att dokumentera det svenska teckenförrådet och att utföra kontinuerlig uppdatering av teckenspråkslexikonet och andra databaser på internet. Redan 1988 började avdelningen planera en lexikografisk databas, konstruktionen av denna startade 1990. Databasen, med filmer med demotecken och meningar, har tillsammans med det första digitala lexikonet på internet, *Digital version av Svenskt teckenspråkslexikon*, vars bilder och teckenbeskrivningar är hämtade från det tryckta teckenspråkslexikonet som skapats av Sveriges Dövas Riksförbund (1997), varit tillgängligt on-line sedan 2001. Sedan 2008 arbetar Avdelningen för teckenspråk med huvuddatabasen *Svenskt teckenspråkslexikon* som finns tillgängligt på institutionens hemsida.[21] Totalt består lexikonet av ca 14 000 tecken och 2 190 exempelmeningar (november 2015).

[21] http://teckenspraklexikon.su.se/

Korpusprojektet tillför *Svenskt teckenspråkslexikon* information om hur tecken används i olika kontexter och ger exempel på meningar där tecknet används. Ett argument för att arbeta parallellt med korpus och lexikon är att korpusannotatörerna behöver lämpliga glosor för tecknen, se Fig. 25 och 26. Målet är att på sikt kunna använda funktionen "kontrollerat ordförråd" (CV) i ELAN för att göra annoteringsarbetet lättare och stabilare. Att glossa korpusmaterial dessutom ger många idéer om vilka tecken som ska inkluderas i lexikonet. Andra områden för det framtida samarbetet inkluderar ett verktyg för att skapa en koppling mellan teckenposter i lexikonet och användningsexempel i korpusen. I andra länder har arbetet för korpusbaserade teckenspråkslexikon påbörjats, bland annat Auslan[22] och BSL.[23]

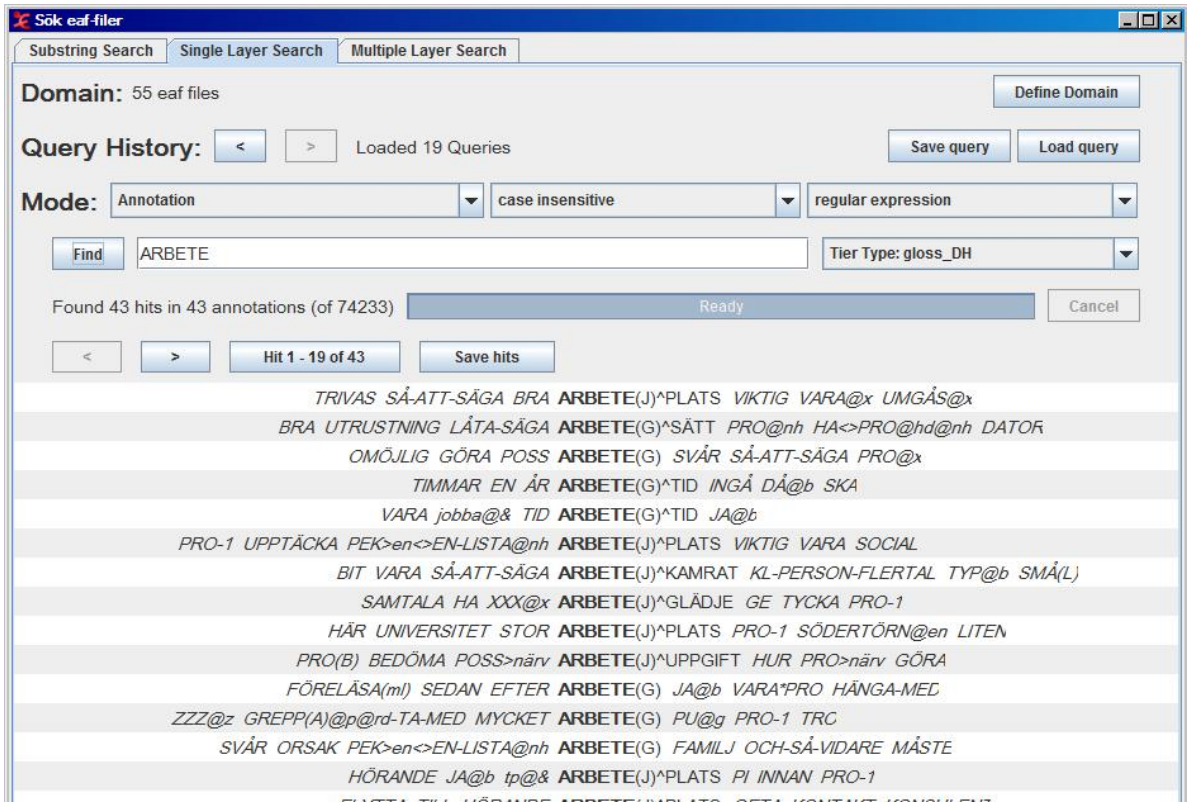
Sammanfattningsvis används korpusundersökningar för teckenspråkslexikografi för att hitta autentiska exempel på tecknets användning, frekvens (hur ofta ett tecken förekommer), ytterligare information om genre, register, kön eller ålder vid tecknets användning, kollokation (semantisk preferens) och kärnteckenförråd för teckenspråksinlärare. För det svenska lexikografiska teckenspråksarbetet är den här korpusen ett utomordentligt viktigt redskap.



Figur 25. Teckentranskription och glosa för "ARBETE" i Svenskt teckenspråkslexikon.

[22] http://www.auslan.org.au/

[23] http://bslsignbank.ucl.ac.uk



Figur 26. Sökning efter glosa "ARBETE" i Svensk teckenspråkskorpus med ELAN. Tilläggen (G) och (J) anger vilket tecken med betydelsen 'arbete' som används.

5.2. Korpusbaserade studier i svenskt teckenspråk

Behovet av tillgång till teckenspråkskorpusar är stort inom teckenspråkslingvistiken. Vid korpusstudier behövs det oftast en stor mängd data för att kunna dra slutsatser utifrån strukturer och för att på statistiska grunder bedöma ett synligt mönster. Forskaren låter datorn sköta jobbet, dvs. den söker och sorterar data på ett objektivt sätt. Korpusstudier är ofta kvantitativa men är även kompletterade med kvalitativa inslag. Forskaren/läraren kan titta på hur ett specifikt tecken i korpusen används i sin aktuella kontext (se McEnery & Xiao 2011). Korpusstudier i teckenspråk är fortfarande mer kvalitativa än korpusstudier i skrivna språk, på grund av tekniska begränsningar (jf korpusar för skrivet språk där bokstäver är lätt maskinläsbara) och korpusstorlek (flera miljoner ord finns i t.ex. Språkbankens korpusar[24]).

Korpusen kan t.ex. också användas i kontrastiv forskning om tvåspråkighet (dvs. svenskt teckenspråk och svenska som dövas andra språk), för tolknings- och översättningsstudier, typologiska studier och jämförelser med andra teckenspråk. Flera studenter i teckenspråkskurser vid Stockholms universitet har använt sig av det svenska korpusmaterialet för sina uppsatser (t.ex. Börsell 2011, Thofelt 2011, Mårtensson 2013, Fallkvist 2014, Bäckström 2015, Ivarsson 2015,

[24] http://spraakbanken.gu.se/

Östling 2015).[25] Jämförelsestudier med flera teckenspråk har genomförts och kommer att genomföras med den del av korpusmaterialet som har samma materialuppläggning, t.ex. fabelberättelser som i ECHO-projektet, eliciteringsmaterial och korta intervjuer, samtal etc. (Crasborn et al, 2008; Jantunen et al, under arbete). Korpusanalys är en växande språkvetenskaplig analysmetod i teckenspråksforskningen (Fenlon, Schembri, Johnston & Cormier, 2015).

En pilotstudie av taggning av teckenspråkstexter genomfördes under hösten 2008 (Wallin & Mesch, 2008) med avsikt att finna en lämplig form av taggning för att underlätta datamaskinell sökning av information ur teckenspråkstexter, så att det går att söka efter ett tecken, ett morfem, en fras, en syntaktisk funktion eller en ordkategori. Pilotstudien visade att det krävs ett omfattande projekt för att sådan taggning ska kunna byggas upp.

Information om tecknens ordklasstillhörighet saknas i det svenska teckenspråkslexikonet. Det är nämligen ingen enkel uppgift att bestämma om ett tecken till exempel hör till ordklassen adjektiv. Man kan använda tre olika tillvägagångssätt för att bestämma ett ords eller teckens ordklasstillhörighet, det bästa sättet för svenskt teckenspråk är att studera tecknets *syntaktiska egenskaper*. Att titta på tecknets *semantiska innehåll*, som är det andra sättet att bestämma ordklasstillhörighet på, kan ibland bli missvisande eftersom det finns tecken som semantiskt visar adjektiviska egenskaper men som syntaktiskt inte kan förekomma som bestämning till substantiv, vilket är ett kriterium för ordklassen adjektiv. Det tredje sättet att bestämma ordklasstillhörighet på är att titta på tecknets morfologi, men detta är det minst tillämpliga sättet för svenskt teckenspråk eftersom språket är uppbyggt på så annorlunda sätt än talade språk. Expriment med ordklassannoteringar som har förts över till svensk teckenspråkskorpus har genomförts i Östlings doktorsavhandling om algoritm (Östling, 2015) och arbetet med ordklasstaggning för svensk teckenspråkskorpus har påbörjats (Börstell et al, under arbete).

För att möjliggöra studier av teckens syntaktiska egenskaper, vilket kan leda till ordklassbestämning av tecken på sikt, fordras ett bra taggningssystem i korpusen. Med detta taggningssystem skulle en lista kunna upprättas som visar var i meningar olika tecken vanligtvis dyker upp, som till exempel om ett tecken vanligen fungerar som bestämning till substantiv eller typiskt förekommer i predikatposition. Det finns internationellt sett fortfarande idag endast få artiklar och beskrivningar som behandlar dessa frågor, men ett exempel är en bruksanvisning för annotering av det australiensiska teckenspråket av Johnston (2010).

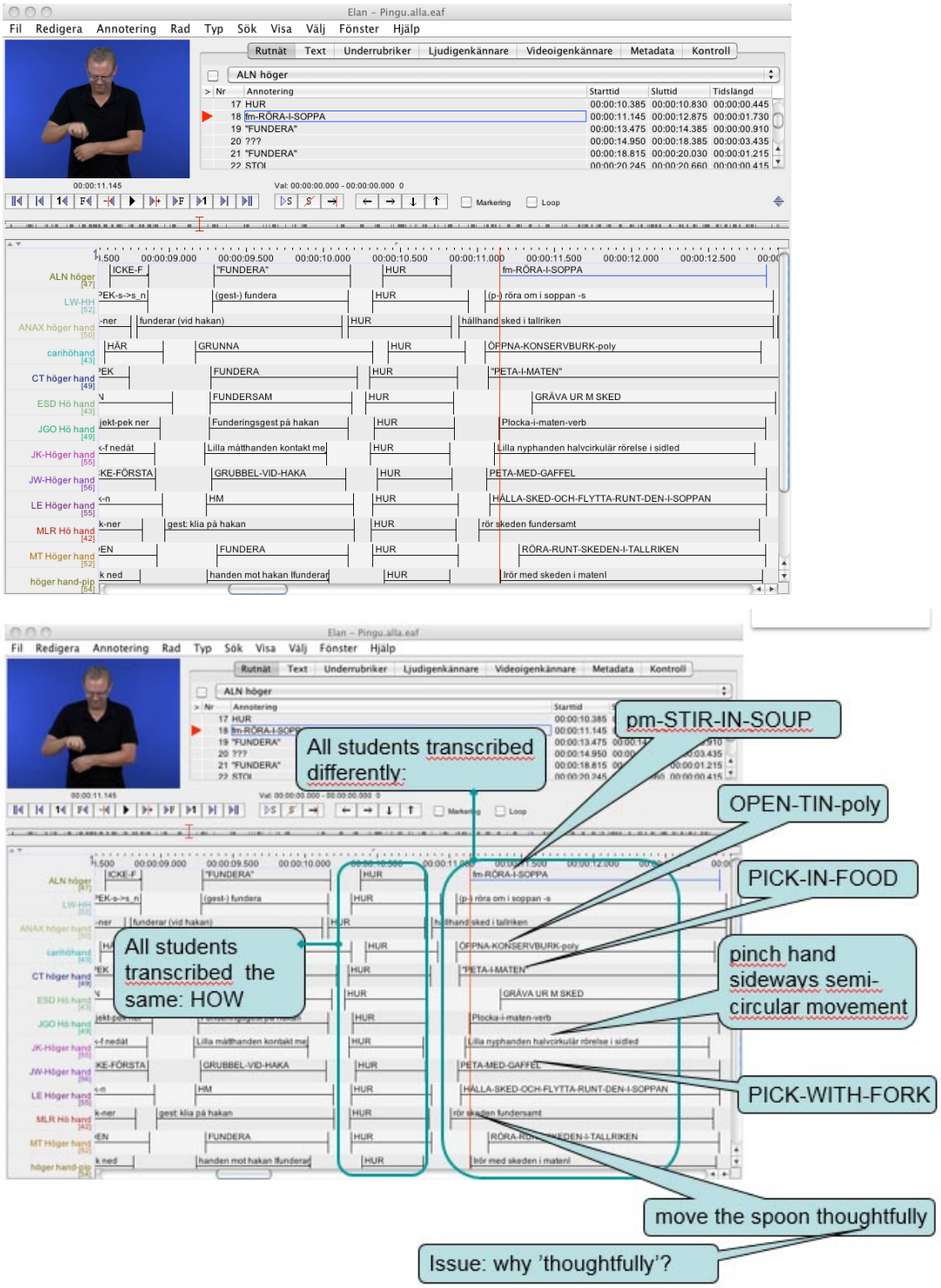
5.3. Undervisningsmaterial

Korpusen är tillgänglig i första hand för lärare och forskare på Avdelningen för teckenspråk vid Institutionen för lingvistik på Stockholms universitet. I filsamlingen finns korpusmaterial, både annoterat och oannoterat, samt tidigare arkivmaterial. Korpusmaterialet används i undervisning och forskning och en del av materialet finns tillgängligt på nätet. I korpusen kan man till exempel söka på olika uttryck, vilket är ett utmärkt verktyg för lärare och inlärare som vill hitta naturliga kontexter för tecken och få se uttrycken användas i autentiskt språkbruk.

Annoteringsarbetet och användningen av ELAN har också intergrerats i teckenspråkskurser vid Stockholms universitet. Här får studenter bland annat läsa om tecknets form och betydelse med hjälp av aspektrader i ELAN, och inlämningsuppgifter i form av eaf-filer kan slås ihop till

[25] *Kan sökas i DiVA www.diva-portal.org*

gemensamma eaf för att skapa överblick över allas svar avseende en annoteringscell, vilken sedan kan diskuteras (se till exempel Fig. 27-28). Ibland genomförs även hemskrivningar som innehåller analysuppgifter med hjälp av ELAN. Korpusmaterialet löser inte alla problem i teckenspråksutbildingen, men det bidrar till utvecklingen av det ämnesdidaktiska arbetet.



Figur 27-28. Ett exempel på tolkstudenters och lärares ihopslagna rader för gruppdiskussion. (En presentationsbild från föreläsningen på workshop i Berlin. Mesch, Nilsson, Wallin & Bäckström, 2010)

6. Slutord

I korpusarbetet har det framgått att annoteringsverktyget ELAN och metadata CMDI bidragit till att skapa en bättre överblick över allt material i databasen. Annoteringskonventioner för svenskt teckenspråk har kunnat utvecklas tack vare det utförda annoteringsarbetet och genom återkommande möten mellan medverkande i korpusarbetet, vilket lett fram till nya insikter om transkriptioner och annoteringsrader, bland annat om rader för tecken som utförs med en respektive två händer i samtalen. Lexikonarbetet i svenskt teckenspråk har också kunnat utnyttjas och det har samtidigt utvecklats parallellt med korpusarbetet.

Korpusprojektet har skapat många möjligheter för framtida teckenspråksforskning. Det har fötts nya ideér till fortsatta studier, annoteringskonventionerna är utvecklingsbara och blir alltmer stabila tack vare det nära samarbetet med lexikonverksamheten. Korpusen är också en stor resurs för svensk teckenspråksforskning, teckenspråksundervisning och tolkutbildning. Sverige har en bra position internationellt sett och har allt sedan starten av korpusarbetet varit aktiva i ett internationellt samarbete i korpusbyggandet. Korpusprojektet hjälper också till att bibehålla den höga position som svensk teckenspråksforskning har haft internationellt sedan 1979, då Sverige arrangerade den första internationella konferensen för teckenspråksforskare.

Det finns dock en del problem i korpusarbetet, främst beroende på begränsade resurser och det omfattande material som ska annoteras, vilket tar lång tid. Dessutom tar redigeringsarbetet mycket tid i anspråk då fem videokameror har använts vid inspelningarna, materialet ska synkas så att man kan hitta samma videosekvens ur olika bildvinklar. Ytterligare en utmaning för korpusprojektet är att underhålla databasen, vilket innefattar hur korpusen ska skötas och hur länge, samt att bevaka att kopieringsrätten (*Creative Commons*) följs. En annan fråga som måste bevakas handlar om i vilka format filmerna ska bevaras för att behålla videoteknisk kvalitet. En kontinuerlig överföring till nya medier är nödvändigt. Institutionen för lingvistik vid Stockholms universitet deltar i projektet Swe-Clarin genom två av sina avdelningar: datorlingvistik och teckenspråk angående digitala resursdatabaser över bland annat teckenspråksdata, med start våren 2015.[26]

[26] http://sweclarin.se/eng/centers/stockholm

Referenser

Ahlgren, Inger & Bergman, Brita. 2006. Det svenska teckenspråket. I *Teckenspråk och teckenspråkiga. Kunskaps- och forskningsöversikt.* SOU 2006:29, sid 11-70. Tillgänglig: http://www.regeringen.se/sb/d/6150/a/60648

Bergman, Brita, Boyes Braem, Penny, Hanke, Thomas & Pizzuto, Elena (Eds.) 2001. Sign transcription and database storage of sign information. Papers from the Intersign network. Special Issue of Sign Language & Linguistics, vol. 4, 1-2.

Borin, Lars. 2011. Resursbehovet för svensk språkteknologi. Föredrag vid seminariet om nationella språkresurser, Stockholm 2011-11-11.

Bäckström, Joel. 2015. Polysyntetiska tecken i svenska teckenspråksdialoger: De vanligast förekommande handformerna i polysyntetiska tecken. Kandidatuppsats i teckenspråk, Institutionen för lingvistik, Stockholms universitet.

Börstell, Carl. 2011. Revisiting Reduplication: Toward a description of reduplication in predicative signs in Swedish Sign Language. Masteruppsats i lingvistik. Institutionen för lingvistik, Stockholms universitet.

Börstell, Carl, Östling, Robert, Mesch, Johanna, Nilsson Björkenstam, Kristina, Sjons, Johan, Wallin, Lars & Wirén, Mats. 2016. Distribution and Duration of Signs and Parts of Speech in Swedish Sign Language. Paper to be presented at the 12th Conference on Theoretical Issues in Sign Language Research (TISLR 12), Melbourne, Australia, 4–7 January, 2016.

Crasborn, Onno. 2004. Technological, linguistic and ethical challenges in sharing sign language data online within the ECHO project. Ms., Radboud University Nijmegen.

Crasborn, Onno. 2010. "What Does "Informed Consent" Mean in the Internet Age? Publishing Sign Language Corpora as Open Content." Sign Language Studies 10:2. 276-290.

Crasborn, Onno, van der Kooij, Els & Mesch, Johanna. 2004. European Cultural Heritage Online (ECHO): publishing sign language data on the internet. Poster presented at the 8th Conference on Theoretical Issues in Sign Lanugage Research, Barcelona, Sept 2004.

Crasborn, Onno, van der Kooij, Els., Waters, David, Woll, Bencie & Mesch, Johanna. 2008. Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language & Linguistics 11:1*, pp. 45-67.

Crasborn, Onno, Mesch, Johanna, Waters, David, Nonhebel, Annika, van der Kooij, Els, Woll, Bencie & Bergman, Brita. 2007. Sharing sign language data online. Experiences from the ECHO project. Internationell *Journal of Corpus Linguistics, 12-4*, pp. 537-564. Tillgänglig: http://www.benjamins.com/cgi-in/t_bookview.cgi?bookid=IJCL%2012%3A4

Fallkvist, Anneli. 2014. Bokstaveringar i svenskt teckenspråk. Bokstaverade lånord och deras ordklasstillhörighet i svenska. Kandidatuppsats i teckenspråk, Institutionen för lingvistik, Stockholms universitet.

Fenlon, Jordan, Schembri, Adam, Johnston, Trevor & Cormier, Kearsy. 2015. Documentary and corpus approaches to sign language research. I E. Orfanidou, B. Woll & G. Morgan (red), *Research Methods in Sign Language Studies: A Practical Guide*, First Edition. John Wiley & Sons, Inc.

Hanke, Thomas. 2007. Sign Language Resources at Hamburg University: iLex & beyond. Föreläsningsanteckningar från ett gästföredrag på Avdelningen för teckenspråk, Institutionen för lingvistik, Stockholms universitet, 21 nov 2007.

Ivarsson, Sofia. 2015. En tankepaus i svenskt teckenspråk: En korpusundersökning av spelande fingrar. Kandidatuppsats i teckenspråk, Institutionen för lingvistik, Stockholms universitet.

Jantunen, Tommi, Mesch, Johanna, Puupponen, Anna & Laaksonen, Jorma (submitted). On the rhythm of head movements in Finnish and Swedish Sign Language sentences.

Johnston, Trevor. 2010. From archive to corpus: transcription and annotation in the creation of signed language corpora. *International Journal of Corpus Linguistics 15(1)*, pp. 104-29.

Johnston, Trevor. 2005. *Corpus of grammar and discourse strategies of deaf native users of Auslan, Australian Sign Language.* Tillgänglig: http://hrelp.org/grants/projects/index.php?land=9

Johnston, Trevor & Schembri, Adam. 2006. Issues in the creation of a digital archive of a signed language. I L. Barwick & N. Thieberger (red), Substainable data from digital fieldwork: Proceedings of the conference held at the University of Sydney, pp. 7-16. 4-6 December 2006. Sydney: Sydney University Press.

McEnery, Tony & Xiao, Richard. 2011. What corpora can offer in language teaching and learning. In Eli Hinkel (ed.), *Handbook of Reseach in Second Language Teachong and Learning: Volume 2*, 364–380. New York, NY: Routledge.

McEnery, Tony, Xiao, Richard & Tono, Yukio. 2006. Corpus-based language studies: An advanced resource book. Taylor & Francis.

Mesch, Johanna. 2011. Att använda ELAN: Bruksanvisning för annotering och studie av teckenspråkstexter. Version 3. Stockholms universitet, Institutionen för lingvistik, Avdelningen för teckenspråk.

Mesch, Johanna, Nilsson, Anna-Lena, Wallin, Lars, Bäckström, Joel. 2010. Using corpus data for teaching purposes. Inbjuden presentation på Sign Linguistics Corpora Network Workshop 4: Exploitation. Berlin, 3 December 2010.

Mesch, Johanna, & Wallin, Lars. 2015. Gloss annotations in the Swedish Sign Language corpus. International Journal of Corpus Linguistics, 20(1), 102-120.

Mårtensson, Katarina 2012. Temabojen i svenskt teckenspråk. Form och användning. Kandidatuppsats i teckenspråk, Institutionen för lingvistik, Stockholms universitet.

Nilsson, Anna-Lena & Rohdell Maya. 2011.”What should I write? Some do´s and don´ts when translating signed corpus material for the web”. Poster presented at the conference of the European Forum of Sign Language Interpreters (efsli), Salemo, Italy. 17 September 2011.

Nonhebel, Annika, Crasborn, Onno & van der Kooij, Els. 2004. Sign language transcription conventions for the ECHO project. URL: http://www.let.kun.nl/sign-lang/echo/docs/transcr_conv.pdf

Thofelt, Unn. 2011. Något om den konstruerade dialogen i svenskt teckenspråk. Kandidatuppsats i teckenspråk, Institutionen för lingvistik, Stockholms universitet.

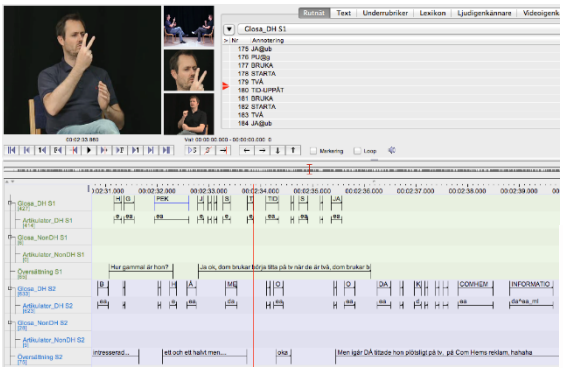
Wallin, Lars & Mesch, Johanna. 2014. Transkriptionskonventioner för teckenspråkstexter. Version 5. Stockholms universitet, Institutionen för lingvistik, Avdelningen för teckenspråk.

Östling, Robert. 2015. Bayesian models for multilingual word alignment. Doktorsavhandling. Institutionen för lingvistik, Stockholms universitet.

Annoteringskonventioner för teckenspråkstexter

Lars Wallin

Johanna Mesch



Innehållsförteckning

Introduktion 30
1.1 Om annoteringskonventionerna 30
1.2 Var börjar och slutar ett tecken? 31
2 Annotering av tecken i Glosa_DH och Glosa_NonDH 32
2.1 Val av ordform till glosa 34
2.1.1 Annoteringsformer 35
2.1.2 Enkla tecken 35
2.1.3 Pluralformer 36
2.1.4 Översättningslån 36
2.1.5 Sammansmälta teckensekvenser 37
2.1.6 Siffertecken 37
2.1.7 Pekningar 37
2.1.8 Possessiva uttryck 39
2.1.9 Objektspronomen 39
2.1.10 Bokstaverade tecken (@b) 40
2.1.11 Klassifikatoriska tecken (@kl) 40
2.1.12 Persontecken och egennamn (@en) 40
2.1.13 Reduplicerade tecken (@rd) 40
2.1.14 Polysyntetiska tecken (@p) 41
2.1.15 Gestlika tecken (@g) 43
2.1.16 Ny förekomst, osäkerhet, rätt glosa?, tecken? (@z) 44
2.1.17 Avbrutet tecken eller tankepaus (@&) 44
2.1.18 Kroppsagerande (@ca) 45
3 Annotering i Artikulator_DH och Artikulator_NonDH 45
4 Annotering av bojar i Glosa_NonDH 47
4.1 Listbojar 47
4.2 Pekboj 48
4.3 Temaboj 48
4.4 Punktboj 49
4.5 AVGRÄNS 49
5 Översättningsrad 50
6 Glosa_DH_extra 51
7 Annotering av icke-manuella komponenter 54

7.1 Mun 54
7.2 Ögon 55
7.2.1 Blickriktning 55
7.2.2 Öppningsgrad 57
7.3 Ögonbryn 58
7.4 Näsa/Kinder 59
7.5 Huvud 60
7.6 Kropp 62
8 Referenser 64
Bilaga 1. Handformssymboler 66
Bilaga 2. Klassifikatoriska tecken 68
Bilaga 3. Polysyntetiska tecken 69

Introduktion

Konventionerna i bruksanvisningen avser annotering av texter med annoteringsverktyget ELAN (se vidare om verktyget i Mesch, 2011 och http://www.lat-mpi.eu/tools/elan/). Syftet är att ge riktlinjer för annoteringsarbete inom projektet ”Korpus för det svenska teckenspråket”, 2009–2011, finansierat av Riksbankens Jubileumsfond (Mesch, Wallin, Nilsson & Bergman, 2012). Konventionerna kan även tjäna som modell för annoteringar utanför korpus, i forskningsprojekt och undervisning.

Att bygga upp en korpus med teckenspråkstexter är ett tidskrävande arbete. Korpusen utgörs inte enbart av annoteringar av tecken utan också av annoteringar av icke-manuella signaler, ordklasser, satsdelar, uppbackningar, fonologisk variation, m.m. Allt kan inte göras i första vändan utan arbetet måste delas upp i flera steg. Det första och viktiga steget är att få alla tecken i korpusen annoterade. Detta är förutsättningen för att göra alla teckenspråkstexter läsbara för datorn, vilket är nyckeln till korpusarbetet. I efterföljande steg kan man sedan komplettera och expandera korpusen med de ytterligare annoteringar som behövs för olika studier (jfr. Johnston, 2015, sid 6).

1.1 Om annoteringskonventionerna

Utgångspunkten för annoteringskonventionerna är att annoteringen så långt som möjligt ska ske på ett sådant sätt att sökning av relevanta data underlättas i ELAN. Verktyget ELAN är konstruerat så att annoterade texter är länkade till videosekvenser i realtid. Sökningar begränsas inte enbart till en enda teckenspråkstext, det går också att göra en sökning i flera texter samtidigt. Då är det viktigt att teckenspråkstexterna har annoterats på ett så enhetligt sätt som möjligt, att samma tecken annoteras med samma glosa oavsett kontext (mer om det i kapitel 2.1).

Annoteringskonventionerna är en vidareutveckling av Bergman (1982): ”Transkription av texter” i *Teckenspråkstranskription* (FOT-rapport X, s. 27-35). En del av konventionerna är också inspirerade av bl.a. *Sign language transcription conventions for the ECHO Project* (Nonhebel et al. (2004) http://sign-lang.ruhosting.nl/echo), *Auslan corpus annotation guidelines* (Johnston (2015)) och CHILDES (MacWhinney (2012) http://childes.psy.cmu.edu/manuals/CHAT.pdf).

Beskrivningen av annoteringskonventionerna är uppdelad i två delar. Första delen beskriver konventioner för annotering av tecken, andra delen beskriver konventioner för annotering av icke-manuella komponenter.

Annoteringskonventionerna är menade att fungera dels som en instruktion för den som ska annotera korpusmaterial, dels som en vägledning i hur man ska läsa de annoteringsfiler som kan laddas ner från korpusens hemsida[27]. Den som annoterar utanför korpusarbetet

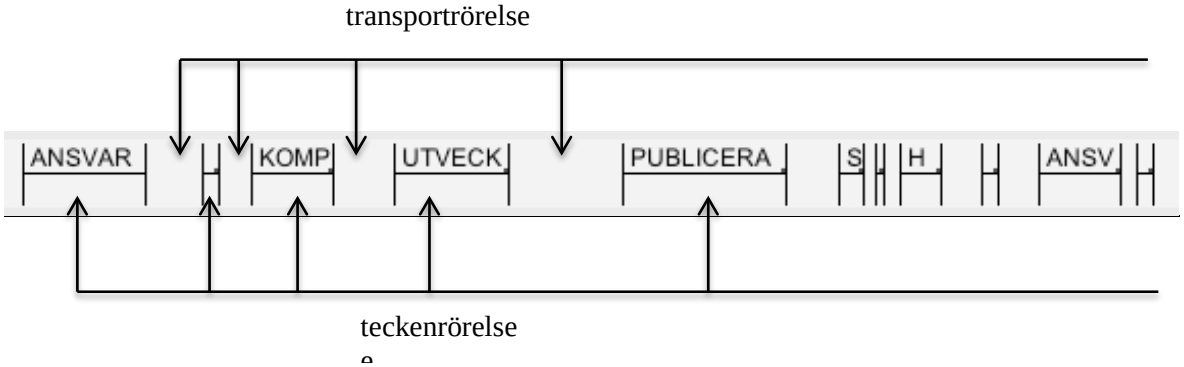
[27] *http://www.ling.su.se/teckenspråksresurser/teckenspråkskorpora/svensk-teckenspråkskorpus*

behöver naturligtvis inte följa konventionerna, men kan förhoppningsvis få inspiration från dem.

1.2 Var börjar och slutar ett tecken?

En teckenspråkstext består av tecken som löper efter varandra. Händerna gör inga stopp mellan tecknen (utom vid pauser eller vid avslut) utan utför kontinuerliga rörelser i varierande tempo allteftersom tecknen produceras. Annotatörens uppgift blir att försöka fastställa vilka delar av rörelsen som utgör tecknens artikulation, dvs. var tecknen börjar och slutar, och vad som inte utgör tecknens artikulation utan istället utgör vad vi kallar för *transportrörelse*, vilket är rörelsen från det att föregående teckens artikulation avslutats och nästföljande teckens artikulation påbörjas. Den kontinuerliga rörelsen består alltså av en kombination av omväxlande teckeninterna rörelser och transportrörelser (se Figur 1).

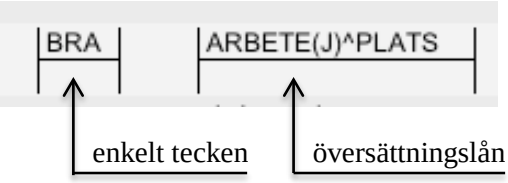
Artikulationernas tidslängd kan variera i olika grad från tecken till tecken, liksom transportrörelsernas tidslängd mellan tecknen. Artikulationens varierande tidslängder hänger samman med de enskilda tecknens fonologiska karaktär – vissa tecken har en kort artikulation medan andra har en längre artikulation. Transportrörelsernas varierande tidslängder hänger samman med var i teckenrummet ett teckens artikulation avslutas och var nästföljande teckens artikulation påbörjas, dvs. om de realiseras nära eller längre bort från varandra.



Figur 1. Kontinuerlig rörelse med omväxlande teckenintern rörelse och transportrörelse

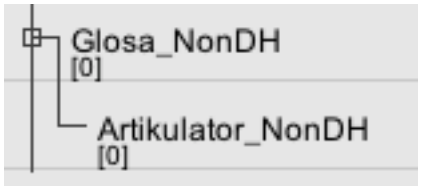
Utöver att fastställa var tecknet börjar och slutar, ska annotatören även ta ställning till om tecknet utgörs av ett enkelt tecken eller om tecknet är en sammansättning av två eller flera tecken i ett översättningslån från svenska. Det sistnämnda fallet innebär att tecknet annoteras med en glosa för varje tecken i sammansättningen, sammanbundna med en cirkumflex (^) mellan glosorna, där annoteringscellen sträcker sig från början av det första tecknets artikulation till slutet av det sista tecknets artikulation (se Figur 2). En cell kan alltså bestå av en eller flera glosor beroende på om tecknet är ett enkelt tecken eller ett översättningslån från ett sammansatt ord.[28]

[28] *Mer om fler glosor i en cell, se kap 2.2.3 och 2.2.4.*



Figur 2. Exempel på ett enkelt tecken och ett översättningslån

2 Annotering av tecken i Glosa_DH och Glosa_NonDH

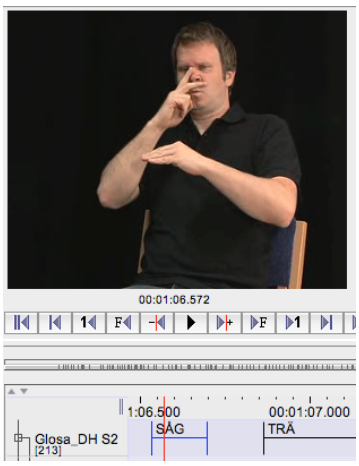


Tecken skrivs företrädesvis i raden Glosa_DH och i förekommande fall även i raden Glosa_NonDH. DH står för dominanta handen, vilket är höger hand hos högerhänta tecknande, respektive vänster hand hos vänsterhänta tecknande. NonDH står för icke-dominanta handen, som är den andra handen (dvs. vänster hand hos högerhänta tecknande och höger hand hos vänsterhänta tecknande).[29]

Hur avgör man vilken hand som är den tecknandes dominanta hand? Det gör man genom att titta på vilken hand som den tecknande normalt använder när denne utför ett tecken med en hand. Ett enhandstecken som SÅG (ID-nr 8752)[30] utförs normalt med den dominanta handen (se Figur 3 och 4). Samma sak gäller för ett tvåhandstecken som SITTA (ID-nr 8200), där vanligtvis den dominanta handen agerar mot den andra handen, som fungerar som manuellt läge (se Figur 5 och 6). I båda fallen där den högerhänta och vänsterhänta tecknande utför tecknen med den dominanta handen, skrivs tecknen i raden Glosa_DH.

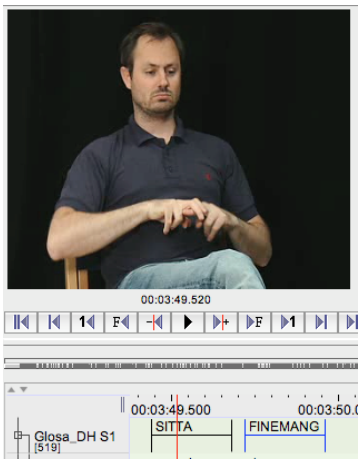
[29] *Det förekommer att en tecknande varken är huvudsakligen högerhänt eller vänsterhänt. Då behöver man för annoteringen bestämma om denne är högerhänt eller vänsterhänt.*

[30] *ID-nr refererar till tecknets nummer i* Svenskt teckenspråkslexikon.

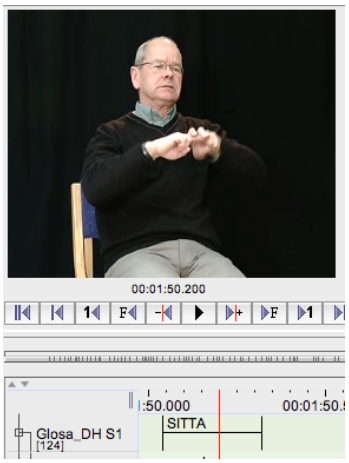


Figur 3. Högerhänt tecknande tecknar SÅG[31]

Figur 4. Vänsterhänt tecknande tecknar SÅG



Figur 5. Högerhänt tecknande tecknar SITTA

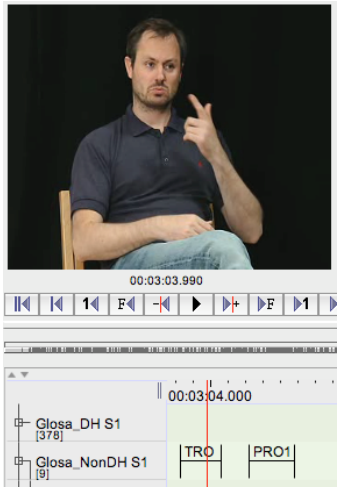


Figur 6. Vänsterhänt tecknande tecknar SITTA

Det kan förekomma att en tecknande växlar hand mitt under teckenproduktionen, så att denne utför ett tecken med den icke-dominanta handen istället för den dominanta handen, dvs. med vänster hand för den högerhänte (se Figur 7) eller höger för den vänsterhänte (se Figur 8). Vid den typen av handväxling skrivs tecknet i raden Glosa_NonDH.

Det kan också förekomma att två tecken utförs samtidigt eller överlappar varandra (se Figur 9). Då skrivs tecknet som utförts med dominanta handen i raden Glosa_DH och tecken som utförts med icke-dominanta handen i raden Glosa_NonDH.

[31] *Fotoillustrationerna är hämtade från Mesch, Rohdell & Wallin, 2012 och 2014.*



Figur 7. Högerhänt tecknande tecknar TRO med icke-dominanta handen, annoterat i raden Glosa_NonDH

Figur 8. Vänsterhänt tecknade tecknar BEGE-SIG-HEM med icke-dominanta handen, annoterat i raden Glosa_NonDH

Figur 9. Dominanta handen utför tecknet STANSOPERATRIS, annoterat i raden Glosa_DH, och icke-dominanta LISTBOJ.EN, annoterat i raden Glosa_NonDH

Sammanfattningsvis skrivs tecken som utförs med den dominanta handen i raden Glosa_DH, och tecken som utförs med den icke-dominanta handen i raden Glosa_NonDH.

2.1 Val av ordform till glosa

Ordet eller orden i Glosa-raderna ska inte ses som en översättning, utan som namn eller etiketter på tecknen. Ord som fångar tecknets grundläggande betydelse har valts. Samma ord eller glosa ska i görligaste mån användas för samma tecken, oavsett resultatet av en översättning till svenska. Exempelvis kan tecknet med glosan FELFRI (ID-nr 443)[32] förekomma i skilda kontexter eller i texter som "alla i släkten är hörande, ingen i släkten är döv", "hemresan gick bra, hände inget på vägen" eller "knäet verkade oskadat, röntgenplåtarna visade inga tecken på fel i knäet". Valet av glosa är inte alltid perfekt. Den första kontexten med hörande och döv har ju inte direkt med felfri att göra, vilket de två andra kontexterna har. Sammantaget blev glosvalet FELFRI det närmaste man kunde komma. Idealet hade naturligtvis varit om teckenspråket hade ett eget ortografiskt skriftsystem. Men det finns inte, så vi får acceptera att använda oss av lån från svenska språket och göra det bästa av det.

Om två tecken med samma form har tydligt skilda betydelser, ska de annoteras med varsin glosa. Exempelvis annoteras tecknen för ’teckenspråkstolk’ (ID-nr 5214) och ’månad’ (ID-nr 5212) med TOLK respektive MÅNAD. Är de betydelsebesläktade men skiljer sig i om

[32] *ID-nr eller teckennummer refererar till tecken i* Svenskt teckenspråkslexikon (http://teckenspraklexikon.su.se).

det ena är substantiv och det andra verb, då annoteras de[33] också med varsin glosa. Till exempel annoteras tecknen för ’jobb’ (ID-nr 4920) och ’jobba’ (ID-nr 2063) med JOBB respektive JOBBA.[34]

Glosorna skrivs med versaler. De ska i görligaste mån skrivas i sin grundform, utan böjningar. Det innebär att glosor för tecken som refererar till entiteter som människor och föremål, skrivs i singular och obestämd form, som STOL (ID-nr 3997), oavsett om det i kontexten handlar om ’en stol’ eller ’flera stolar’. Tecken som betecknar händelser och handlingar skrivs utan tempusböjning, som ÄTA (ID-nr 1267), oavsett om det i kontexten handlar om nutid eller dåtid. Tecken som betecknar en egenskap eller ett tillstånd, motsvarande adjektiv eller adverb, skrivs i grundform som SNABB (liknar FORT ID-nr 3169), oavsett om kontexten handlar om ’snabba bilar’ eller ’springa snabbt’.

Det finns tillfällen där principen om grundform kan frångås. Ett exempel är att en del svenska ord företrädesvis förekommer i pluralform istället för singularform. Exempel på dessa är *föräldrar* och *byxor*. Tecken för dessa begrepp annoteras lika, dvs. FÖRÄLDRAR (ID-nr 5787) och BYXOR (ID-nr 555). Ett annat exempel är tecknet som liknar SE (ID-nr 4104) men som annoteras med preteritumformen SÅG (efter munrörelsen /såg/)[35].

Om man är osäker på vilken ordform eller glosa som ska användas för ett tecken kan man gå till *Svenskt teckenspråkslexikon*, där det finns glosor för de flesta tecknen.[36] Här nedan ges några exempel på hur glosor ska skrivas.

2.1.1 Annoteringsformer

Tecken ska generellt annoteras med versaler. Undantag är tecken som avbryts vid felsägning eller en rörelse som påminner om ett tecken men som inte säkert är det. Dessa fall skrivs istället med gemener (se vidare om gemener i kapitel 2.2.13).

2.1.2 Enkla tecken

Tecken annoteras med svenska ord i versaler:

FRÅGA	(ID-nr 4540)
RÄTT	(ID-nr 387)
TRAFIK	(ID-nr 1037)

[33] *Det pågår en diskussion om att annoteringen inte ska skilja på ordklasser, dvs. att samma teckenform för 'jobb' och 'jobba' gärna ska annoteras gemensamt med lemmaformen JOBBA.*

[34] *Konventionen att annotera efter ordklasstillhörighet är inte självklar. Några menar att man ska annotera med samma glosa oavsett ordklasstillhörighet och följa samma lemmaregel som för FELFRI, dvs. JOBBA väljs oavsett kontexten, oavsett om det handlar om jobb eller jobba.*

[35] *Detta är två skilda tecken, vilket man märker i betydelseskillnaden när de kombineras med tecknet INTE. SÅG*INTE uttrycker att man t.ex. inte såg något hända, dvs. att man inte observerat något. SE*INTE uttrycker att man inte kan se något, t.ex. pga. otillräcklig synförmåga eller att en text är för liten.*

[36] *Svenskt teckenspråkslexikon uppdaterar kontinuerligt glosor till alla tecken.*

Om ett tecken behöver annoteras med en sekvens av två eller flera ord, separeras orden med bindestreck. Bindestreck symboliserar alltså ingen teckensekvens, utan ett enkelt tecken:

FÅ-SYN-PÅ (ID-nr 1497)

KOMMA-DIT(Lb) (ID-nr 2830)

Varje tecken ska så långt som möjligt annoteras med en unik glosa. Det förekommer dock fall, när två olika tecken betecknar samma typ av begrepp, exempelvis finns det två tecken som båda betyder 'ensam'. Det ena är ett tecken som utförs med pekfingerhanden (ID-nr 2862) och det andra med tumhanden (ID-nr 2360). I det fallet annoteras de två tecknen med specificerad handform, tumhand med (B) och pekfingerhand med (L), som i följande exempel:[37]

ENSAM (B) (ID-nr 2360)

ENSAM (L) (ID-nr 2862)

Ett annat exempel är två olika tvåhandstecken som båda betyder 'fortfarande'. Det är inte handformerna som skiljer de båda tecknen åt, skillnaden är att det ena tecknet utförs med båda händerna aktiva, dvs. dubbel artikulator (ID-nr 983), och det andra tecknet utförs med ena handen vid den andra handen, dvs. enkel artikulator med manuellt läge (ID-nr 12317). Dessa och liknande tecken annoteras med ett tillägg till glosan: dubbel artikulator med (da) och enkel artikulator med manuellt läge med (ml):

FORTFARANDE(da) (ID-nr 983)

FORTFARANDE(ml) (ID-nr 12317)

Samma exempel men med andra funktioner är tecknen för 'annan'. Ena tecknet utförs som enhandstecken (ID-nr 4333) och annoteras med (ea) för enkel artikulator (utan manuellt läge). Det andra tecknet utförs som tvåhandstecken med manuellt läge (ID-nr 2994) och annoteras med (ml) som följande:

ANNAN(ea) (ID-nr 4333)

ANNAN(ml) (ID-nr 2994)

2.1.3 Pluralformer

Substantivtecken förekommer normalt i singularis. Bland dessa finns det dock en liten grupp tecken som har en lexikal pluralmotsvarighet. Dessa är BILD, BOK, DAG, FLICKA, GÅNG, KARL, KOMPIS, KVÄLL, LAND, MEDLEM, MÅNAD(J), MÅNAD(L), POJKE, SAK och UPPGIFT. I singularis utförs de med enkel artikulation, men i pluralis med upprepad artikulation. Pluralformerna annoteras med glosa i singularis men med tillägget .PL (förkortning av plural), t.ex. BILD.PL.

2.1.4 Översättningslån

Teckenspråket lånar en del sammansatta ord från det svenska språket. Som exempel uttrycks 'sommarjobb' på teckenspråk med tecknet för 'sommar' SOMMAR (ID-nr 5072)

[37] *Se vidare bilaga 1 med uppsättning symboler som används i annoteringen.*

följt av tecknet för 'jobb' JOBB (ID-nr 4919). För att markera att de tillsammans utgör ett översättningslån annoteras de i samma cell med en cirkumflexsymbol ”^” mellan tecknen, som följer:

SOMMAR^JOBB (ID-nr 5072; 4919)

2.1.5 Sammansmälta teckensekvenser

Det finns andra typer av teckensekvenser som utförs som en sammansmält enhet trots att de inte är översättningslån. I den sammansmälta sekvensen brukar ett eller flera av tecknen utföras något annorlunda än samma tecken i fristående position. Om vi exempelvis tittar på teckenparet FÖRSTÅ (ID-nr 2570) och INTE (ID-nr 646), som tillsammans uttrycker 'förstår inte', utförs dessa i fristående position FÖRSTÅ med pekfingerhand som förs framåt från pannan, och INTE med båda flata händerna som förs åt sidorna med handflatorna vända uppåt framför kroppen. I den sammansmälta sekvensen går tecknens artikulation in i varandra med pekfingerhanden som förs åt sidan från pannan med en vridrörelse och handformen förändras till flat hand i en enda följd (ID-nr 2567). Tecknet för 'förstå' och tecknet för 'inte' finns fortfarande, men i ett annorlunda utförande. Dessa och liknande exempel på sammansmälta sekvenser annoteras i samma cell med asterisksymbol ”*” mellan tecknen som i följande exempel:

FÖRSTÅ*INTE (ID-nr 2567)

2.1.6 Siffertecken

Siffertecken är tecken som uttrycker tal, pengar, tider eller årtal. De annoteras inte med siffror, utan med bokstäver och i förekommande fall som sammansatta översättningslån.

TJUGO^FEM

EN^TUSEN^ÅTTA^HUNDRA

NITTON^HUNDRA^SEXTIO^TRE

2.1.7 Pekningar

Här beskrivs annoteringar av pekningar som noterats i korpusen. Pekningarna har annoterats i tre huvudformer: PRO1, PEK och PEK> som beskrivs närmare nedan.

- PRO1 avser pekning som pekar på den tecknade (jfr ID-nr 2798).[38] Den utförs vanligen med pekfingret men det förekommer också att den utförs med hela handen eller med tummen. De annoteras då med PRO1(J) för hela handen och PRO1(B) för tummen.

Vid flertalsuttryck annoteras PRO1 på ytterligare tre olika sätt. Det första är PRO1.FL. Pekfingerhanden gör en svepande rörelse vid den tecknades kropp (jfr ID nr 5491 och 8535). Det andra är PRO1.DUAL. Här består handen av pek- och långfinger som rör sig mellan den tecknade och en punkt i rummet (jfr ID-nr 5490). Det tredje är

[38] *PRO1 står för första person pronomen. Svenskt teckenspråk beskriver personliga (och possessiva) pronomina som ett två-pronomen-system: första person och icke-första person. (Mer om pronomen-systemet kan t.ex. läsas i Nilsson, 2007.)*

PRO1(W).MULTI och PRO1(4).MULTI. Den innefattar att handen, lätt framåtriktad, med pek-, lång- och ringfingret (W) eller alla fingrar utom tummen (4), rör sig i en cirkel i närheten av den tecknades kropp.

- PEK avser pekningar som är riktade från den tecknade. Den utförs vanligen med pekfingret men det förekommer också att den utförs med hela handen eller med tummen. De annoteras då med PEK(J) respektive PEK(B).

Vid flertaluttryck annoteras PEK med någon av variationerna av de tre olika sätten som beskrivs för PRO1 men med skillnad i artikulation. Vid PEK.FL gör handen med pekfingret en svepande rörelse i rummet framför den tecknades kropp. Vid PEK.DUAL gör handen med pek- och långfinger en sidorörelse i rummet framför den tecknades kropp (jfr ID nr 11454). Vid PEK.MULTI gör handen inte en cirkelrörelse, istället upprepas pekningar i olika riktningar.

- PEK> avser pekningar som pekar på någon kroppsdel eller personer eller entiteter som är närvarande eller synliga i filmen. Den annoteras följt av person (PEK>person), typ av entitet (PEK>bok) eller en kroppsdel (PEK>knä).

Det finns en variant för flertal – PEK.FL>. Då utförs en svepande rörelse som inkluderar närvarande referenter. Om det är personer blir det PEK.FL>person.

Vanligaste handformen är pekfingret. Andra handformer är hela handen (PEK(J)>), pek- och långfinger (PEK(V)> eller pek-, lång- och ringfinger (PEK(W)>).

Som angetts inledningsvis annoteras det som utpekas. Person anges naturligt med person. Kroppens yttre och inre delar anges med deras namn, som axel och nyckelben.[39] Fingrarna annoteras som följande: pekf, långf, ringf, lillf och tumf.

För att sammanfatta:

PRO1	PRO1(J)	PRO1.FL	PRO1.DUAL	PRO1(W).MULTI
	PRO1(B)			PRO1(4).MULTI
PEK	PEK(J)	PEK.FL	PEK.DUAL	PEK.MULTI
	PEK(B)			
PEK>	PEK(J)>	PEK.FL>		
	PEK(V)>			
	PEK(W)>			

Det finns pekande tecken som inte annoteras med PEK utan med andra glosor. Exempel på dessa är UPP (ID-nr 8529), NER (ID-nr 5109), HÄR (ID-nr 2866) eller DÄR (4726).

[39] *Observera att man inte ska blanda ihop pekningarna med lexikala deiktiska tecken för 'ögon', 'hjärta', 'näsa' etc. De annoteras med glosorna ÖGON, HJÄRTA och NÄSA.*

2.1.8 Possessiva uttryck

Tecken med possessiv betydelse liknar pekningar fast med hela handen och handflatan vänd mot rörelseriktningen (jfr ID-nr 187 och 275). Tecknen annoteras förkortat med POSS. Handflatan vänd mot den tecknade annoteras med POSS1, mot synlig referent med POSS> och utåt mot rummet med POSS.

Vid flertalsuttryck förekommer POSS1.FL med svepande rörelse vid den tecknades kropp (jfr ID-nr 186), POSS.FL med svepande rörelse i rummet framför den tecknades kropp (jfr ID-nr 2140), och POSS.MULTI med upprepade rörelser i olika rörelseriktningar.

Det finns en annotering som är förbehållen POSS. Det är POSS.EGEN> och POSS.EGEN. EGEN är en förkortning av egenskap. Tecknet syftar på en referents egenskap, karaktär eller egenhet. Tecknet kombineras ofta med särskild munaktivitet som liknar /bi/, dvs läpparna går från stängd till öppen.

För att sammanfatta:

POSS1	POSS1.FL	
POSS	POSS.FL	POSS.MULTI
POSS>	POSS.FL>	
POSS.EGEN		
POSS.EGEN>		

2.1.9 Objektspronomen

Objektspronomen har liksom POSS handflatan vänd utåt eller inåt fast med annan handform och artikulation. Handformen är böjda tummen och pekfingret samt rörelseriktning ner vid entalsuttryck. Tecken med handflatan vänd mot den tecknade annoteras med OBJPRO1 (jfr ID-nr 3622), mot synlig referent med OBJPRO> och utåt mot rummet med OBJPRO (jfr ID-nr 3639).

Vid flertalsuttryck förekommer OBJPRO1(Lbt).FL och OBJPRO1(L).FL med svepande rörelse vid den tecknades kropp (jfr ID-nr 4134, 5147 och 8488) och OBJPRO.FL med svepande rörelse framför den tecknades kropp.

För att sammanfatta:

OBJPRO1	OBJPRO1(Lbt).FL
	OBJPRO(L).FL
OBJPRO	PEK.FL
OBJPRO>	PEK.FL>

2.1.10 Bokstaverade tecken (@b)

Tecken som består av bokstavering annoteras med @b[40] efter glosan enligt följande:

ALBINO@b

LIRARE@b

2.1.11 Klassifikatoriska tecken (@kl)

Det finns en grupp tecken som beskrivs som *noun classifiers* (Bergman & Wallin, 2001). Här nämns de som klassifikatoriska tecken. Det är en grupp tecken som typiskt förekommer direkt efter substantiv. Deras kännetecken är att teckenutförandet, ofta men inte alltid, är reducerat i jämförelse med motsvarande lexikala tecken och att munrörelsen är en fortsättning från föregående substantiv. Tecken som annoterats som klassifikatoriska tecken visas i Bilaga 2.

2.1.12 Persontecken och egennamn (@en)

Persontecken annoteras med personens fulla namn. Namntecken annoteras med namnet.[41] Samma gäller tecken för egennamn på organisationer, företag, fordon, länder, städer, etc. Glosan följs av kategorimarkören @en (förkortning av *egennamn*). Några exempel:

LARS-WALLIN@en

ANDERS@en

HEPHATA@en

Skulle tecken för egennamn utgöras av bokstavering, skrivs @b direkt efter glosan följt av @en som följande exempel:

ÖSK@b@en

SAAB@b@en

2.1.13 Reduplicerade tecken (@rd)

Det förekommer att tecken utförs i reduplicerad form, dvs. att tecknets rörelsecykel upprepas ett antal gånger.[42] Detta ska inte blandas ihop med tecken som har upprepning i sin grundform. Reduplicerade tecken annoteras med glosa (efter grundtecknet) följt av kategorimarkören @rd (förkortning av *reduplikation*). Några exempel:

SKRIVA@rd (SKRIVA ID-nr 3410)

HÄNDA@rd (HÄNDA ID-nr 3023)

---

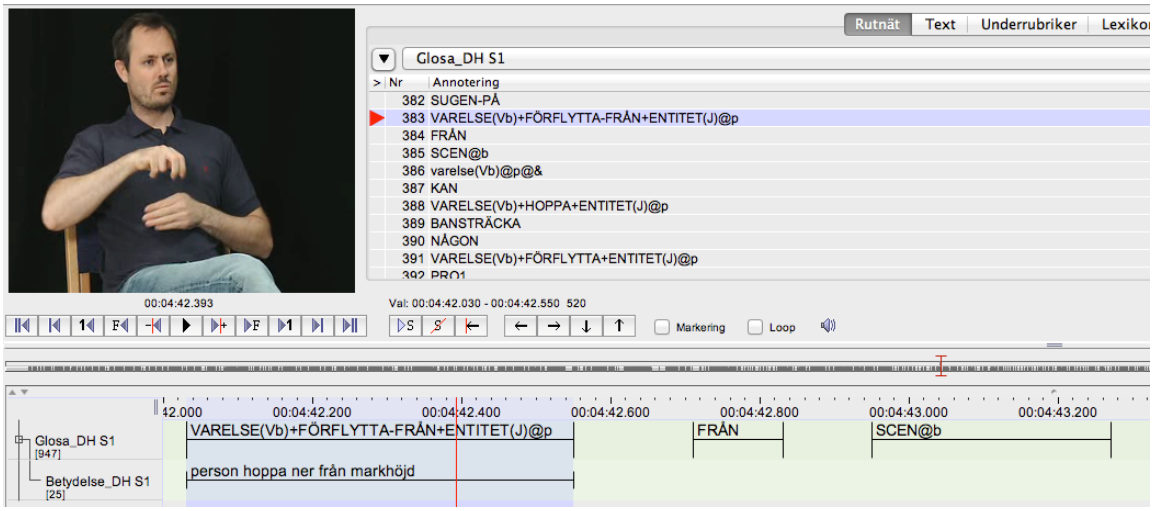
[40] *@ är en symbol som används som tag för olika kategorier.*

[41] *Mer om persontecken och namntecken kan läsas i Hedberg (1989).*

[42] *Mer om reduplikationer kan läsas i Bergman (1990), Bergman & Dahl (1994) och Börstell (2011).*

2.1.14 Polysyntetiska tecken (@p)

Polysyntetiska tecken (Wallin, 1994)[43] utgör en grupp tecken med komplicerat innehåll som avbildar handlingar, händelser och tillstånd. Det inkluderar också tecken som beskriver eller avbildar entiteters form och storlek. Att annotera polysyntetiska tecken är mer komplicerat än att annotera lexikala tecken. Konventionen är ännu inte helt färdig, utan är fortfarande under bearbetning. Figur 10 visar ett exempel på hur annoteringen av ett polysyntetiskt tecken kan se ut.



Figur 10. Annoteringen av polysyntetiskt tecken

Tecknet i figur 10 avbildar en situation där en person hoppar ner från en scen (dominanta handens dubbelkrokhand (Vb) förs från icke-dominanta handens flata hand (J) ner framför flata handen).[44] Det annoteras inte med PERSON-HOPPA-NER-FRÅN-SCEN som man kunde förvänta sig. Anledningen till detta är att samma tecken kan beskriva en annan situation, såsom 'katt hoppar ner från en trappavsats'. Istället för att annotera enligt innehållet i den aktuella kontexten, annoterar vi enligt semantiska egenskaper och typ av situation. Person och katt har den gemensamma egenskapen att vara varelser med ben och annoteras med glosan VARELSE. Scen och trappavsats är exempel på föremål eller underlag. De annoteras med glosan ENTITET som också inkluderar kroppsdelar (ej hand). Vätskor och liknande annoteras med VÄTSKA.

Kvar återstår då annotering av den typ av situation som 'hoppa ner från' är exempel på. Andra liknande situationstyper kan vara gå, springa, falla. Gemensamt är att någon eller något förflyttar sig (egenförflyttning) från en plats till en annan plats och annoteras därför med FÖRFLYTTA, i exemplet i Figur 10, FÖRFLYTTA-FRÅN. Tecknet i Figur 10 annoteras enligt följande struktur med VARELSE(Vb)+FÖRFLYTTA-

---

[43] *Teckenkategorin kallas* avbildade tecken *i Bergman (2012) och* avbildande verbkonstruktioner *i Simper-Allen (2013).*

[44] *Mesch, Wallin, Nilsson & Bergman (2012): SSLC01_007 (00.04:42.030-00.04:42.550).*

FRÅN+ENTITET(J)@p. Annoteringen '+' markerar att glosorna betecknar semantiska delar och @p är förkortning för polysyntetiska tecken.

Samtidigt med annoteringen av polysyntetiska tecken beskrivs dess innehåll i en underordnad rad, Betydelse_DH eller Betydelse_NonDH, efter vilken hand som utfört tecknet.

Alla polysyntetiska tecken annoteras inte med en sekvens av tre glosor. En del annoteras med två och resten med en. När icke-dominanta handen fungerar som manuellt läge (som i figur 10), annoteras tecknet med tre glosor, varav den första i sekvensen representerar dominanta handen, den andra artikulationen och den tredje det manuella läget. När tecknet utgörs av dubbel artikulator (dvs. att icke-dominanta handen fungerar som artikulator tillsammans med dominanta handen) annoteras tecknet med två glosor med dubbla handformssymboler, som t.ex. ENTITET(JJ)+FÖRFLYTTA-TILL(@p. Det förekommer fall när polysyntetiska tecken produceras utan markerad artikulation eller med utebliven artikulation. Då annoteras tecknet med en glosa, som t.ex. ENTITET(J)+@p.

VARELSE och ENTITET utgör två av flera handformskategorier. Varje kategori kan representeras av olika handformer, således utgör VARELSE (Vb) och ENTITET (J) en av flera handformer i respektive kategori. Andra handformskategorier som används i korpusen är HAND, GREPP, FORM, VÄTSKA OCH KOLLEKTIV.

VARELSE används vid tecken som uttrycker:

a) befintlighet eller existens, som t.ex. stå eller finnas, annoteras med BEFINNA,

b) förflyttning (som beskrivs ovan), annoteras med FÖRFLYTTA

c) rörelse (utan förflyttning), som t.ex. sträcka på sig eller stå på händer, annoteras med RÖRELSE

d) håller sig stilla (utan rörelse), annoteras med STILLASTÅENDE

ENTITET används vid tecken som uttrycker:

a) befintlighet eller existens, som t.ex. stå eller finnas, annoteras med BEFINNA,

b) förflyttning (som beskrivs ovan), annoteras med FÖRFLYTTA

c) rörelse, som t.ex. lyfta på ben, annoteras med RÖRELSE

d) varken rör eller flyttar på sig, dvs. håller sig stilla, annoteras med STILLASTÅENDE

HAND används vid tecken som uttrycker:

hantering av ett föremål, t.ex. hålla en tidning eller kasta en boll, annoteras med HANTERA

GREPP används vid tecken som uttrycker:

att utföra en handling med hantering av ett verktyg (ej hand), som t.ex. slå tennisboll (med tennisracket), bita någon på näsan (med käften), annoteras med HANTERA.

VÄTSKA används vid tecken som uttrycker:

vatten som rinner eller eld som sprutar ur svets, och annoteras med FÖRFLYTTA.

KOLLEKTIV används i tecken som uttrycker:

en grupp eller massa som förflyttar sig, och annoteras med FÖRFLYTTA.

FORM används vid tecken som beskriver form och storlek och kombineras med:

a) UTSTRÄCKNING när tecknet beskriver en entitet med *en* dominerande utsträckning

b) BESKRIVNING när tecknet beskriver en entitet med mer eller mindre varierad utsträckning som snögubbe eller slingande väg.

För att sammanfatta:

Tecken med manuellt läge	GLOSA(hf)+GLOSA(situation)+GLOSA(hf)@p
Tecken med dubbel artikulator	GLOSA(hfhf)+GLOSA(situation)@p
Tecken utan artikulation	GLOSA(hf)+@p eller GLOSA(hfhf)+@p
GLOSA(hf):	**GLOSA(situation):**
ENTITET och VARELSE	BEFINNA, RÖRELSE, FÖRFLYTTA, STILLASTÅENDE
GREPP och HAND	HANTERA
FORM	UTSTRÄCKNING, BESKRIVNING
VÄTSKA och KOLLEKTIV	FÖRFLYTTA

2.1.15 Gestlika tecken (@g)

Det förekommer ett par återkommande teckenutföranden som ännu inte är lexikalt bestämda men som fyller en väsentlig roll i texten. Ett av dem utgörs av handen/händerna hålls framför kroppen med handflatan eller handflatorna vanligtvis vända uppåt (Figur 11). Detta annoteras tills vidare som ett gestlikt tecken med glosan PU (en förkortning av engelskans ”palms up”) följt av kategorimarkören @g (förkortning för *gest*).[45] Ett annat teckenutförande utgörs av båda händerna hålls på huvudnivå, ovanför axlarna, med

[45] *Den engelska användningen har valts för att underlätta internationellt samarbete. Ett liknande tecken, med likartad funktion/betydelse, verkar finnas i flera tecknade språk (se exv. Roush, 2007).*

fingrarna riktade uppåt och handflatorna vända framåt (Figur 12). Det annoteras med glosan PF (en förkortning av engelska ”palms forward”) följt av kategorimarkören @g.

Figur 11. Exempel på PU@g

Figur 12. Exempel på PF@g

2.1.16 Ny förekomst, osäkerhet, rätt glosa?, tecken? (@z)

Vid ny förekomst av tecken, dvs. tecken som inte tidigare har dykt upp i annoterade texter och som nu dyker upp första gången, annoteras tecknet med GLOSA: följt av förslag på glosa och @z, t.ex. GLOSA:TOMAT@z. Om man inte kommer på ett förslag till glosa kan man välja en glosa från en tidigare förekomst av tecken som tecknet i fråga påminner mycket om och sätta inom parentes, t.ex. GLOSA:(JÄSA)@z. Fungerar inte någotdera, annoterar man med ett '?' inom parentes, t.ex. GLOSA:(?)@z. Om man är osäker på om man uppfattat tecknet rätt eller valt rätt glosa, kan man alltid lägga till markören '@z' efter glosan, t.ex. BARA(B)@z. Allt förutsätter att man är säker på att det är ett tecken informanten utfört. Om det bara påminner om ett tecken utan att vara det, kan man välja att annotera med ZZZ@z.

Markören '@z' är tänkt att informera att annoteringen behöver bearbetas, att den nya glosan ska bekräftas eller behöver ses över ytterligare för att bestämma och/eller bekräfta val av glosa, eller om det verkligen förekommer ett tecken. När bearbetningen är tillfredställande utförd, tas markören bort.

2.1.17 Avbrutet tecken eller tankepaus (@&)

Ibland fullbordar inte den tecknade ett tecken utan avbryter det, t.ex. vid felsägning. Då annoteras det med en glosa i gemener (om man vet tecknet), t.ex. annan@& eller med glosa@& (om man inte kan uttyda tecknet).

Det förekommer att den tecknade stannar upp mitt i teckenutförandet eller gör tvekande avbrott, ibland med spelande fingrar, för att fundera eller söka efter nästa tecken. Då annoteras avbrottet med tp@& (förkortning av tankepaus).

2.1.18 Kroppsagerande (@ca)

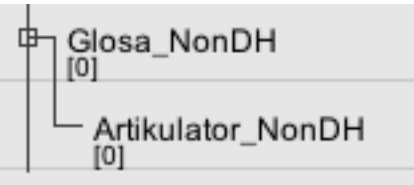
Med kroppsagerande (kroppsrörelse) avses när den tecknade imiterar ett beteende eller agerande, t.ex. när man ropar (Figur 13), agerar målvakt (Figur 14) eller poserar (Figur 15). De påminner om vad forskningslitteraturen kallar *constructed action*, eller förkortat CA, (t.ex. Metzger 1995). Den typen av beteende eller agerande annoterar man med beskrivning följt av @ca som visas under Figur 13, 14 och 15.

Figur 13. ROPA@ca

Figur 14. MÅLVAKT@ca

Figur 15. POSERA@ca

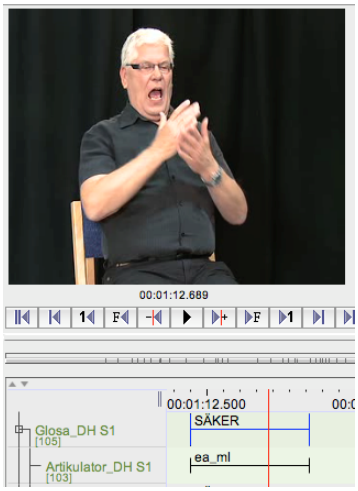
3 Annotering i Artikulator_DH och Artikulator_NonDH



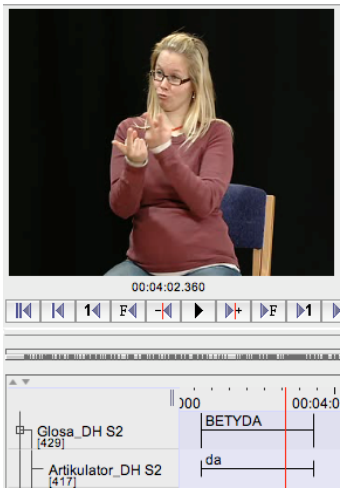
Raderna Artikulator_DH och Artikulator_NonDH utgör en underordnad rad[46] till Glosa_DH respektive Glosa_NonDH. Här annoteras information om vilken typ av tecken som utförts. Tecken som utförts med en hand vid olika lägen utom vid andra handen, som t.ex. BRA (ID-nr 85) annoteras som enkel artikulator eller förkortat ”ea” (Figur 16). Tecken som utförts med en hand vid andra handen, som t.ex. SÄKER (ID-nr 388), annoteras som enkel artikulator med manuellt läge, eller förkortat ”ea_ml” (Figur 17). Tecken som utförts med båda händerna aktiva, som t.ex. BETYDA (ID-nr 8226), annoteras som dubbel artikulator, eller förkortat ”da” (Figur 18).

[46] *Man skiljer på* överordnade *eller* oberoende *rader (på engelska* parent tier*) och* underordnade *eller* beroende *rader (på engelska* child tier*)*

Figur 16. Enkel artikulator



Figur 17. Enkel artikulator med manuellt läge



Figur 18. Dubbel artikulator

När ett tecken annoteras med en cirkumflex eller en asterisk i Glosa-raderna, annoteras också båda symbolerna i Artikulator-raderna, som i Figur 19 och 20. I Figur 19 anger annoteringen att första tecknet i det sammansatta tecknet utförts med enkel artikulator (utan manuellt läge) och andra tecknet med dubbel artikulator. I Figur 20 anges att båda tecken utförs med enkel artikulator (utan manuellt läge).

Figur 19. Sammansatt tecken

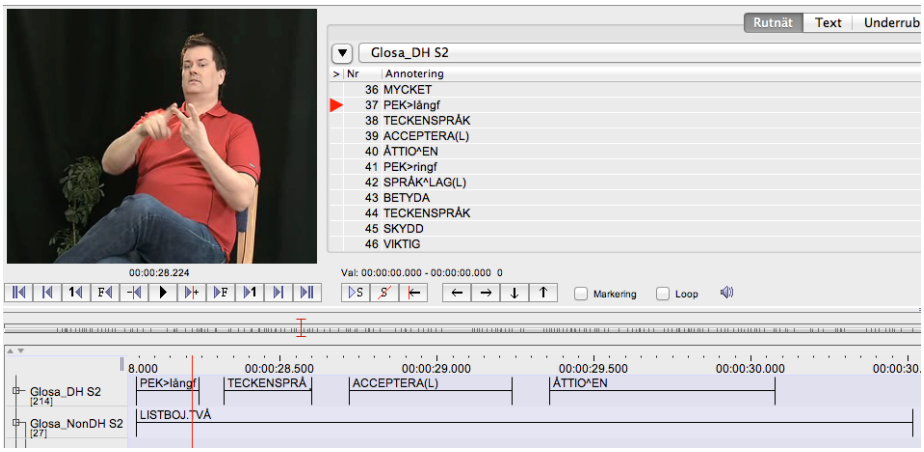
Figur 20. Sammansmält tecken

4 Annotering av bojar i Glosa_NonDH

Nedan listas ett antal tecken, s.k. bojar, som huvudsakligen produceras med den icke-dominanta handen. Gemensamt för dem är att de typiskt hålls i stationär position medan den dominanta handen utför tecken.

4.1 Listbojar

Figur 21 visar ett exempel på en listboj (från eng. ”list buoy”) (jfr. Liddell 2003; Liddell, Vogt-Svendsen & Bergman, 2007; Nilsson, 2007) utförd med den icke-dominanta handen (vänstra) med en handform med pekfinger och långfinger åtskilda. Handformen liknar ett siffertecken för grundtalet ’två’. Karaktäristiskt för listbojarna är att de antar handformer som liknar siffertecken för grundtal, men med handen mer lutad och framför kroppen.



Figur 21. Exempel på listboj, annoterad LISTBOJ.TVÅ i NonDH-raden

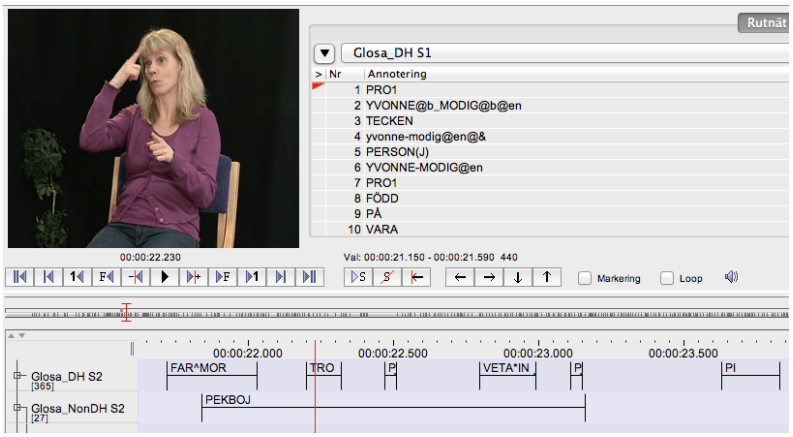
Då listbojen liknar siffertecknet annoteras därför listbojen med grundtalen ett till fem enligt följande:

LISTBOJ.EN	(pekfinger)
LISTBOJ.TVÅ	(pek- och långfinger)
LISTBOJ.TRE	(pek-, lång- och ringfinger)
LISTBOJ.FYRA	(alla fingrar exkl. tummen)
LISTBOJ.FEM	(alla fingrar inkl. tummen)

Det är vanligt att den dominanta handen riktas mot listbojens fingrar. I Figur 21 riktas dominanta handens PEK mot listbojens långfinger. Det utförandet annoteras med en riktningsangivelse ”>” efter glosan PEK följt av ett förkortat namn på fingret, dvs. långf: PEK>långf (se rad 36 efter det röda märket i Figur 21). Varje finger annoteras med fingernamn: pekf, långf, ringf, lillf och tumf.

4.2 Pekboj

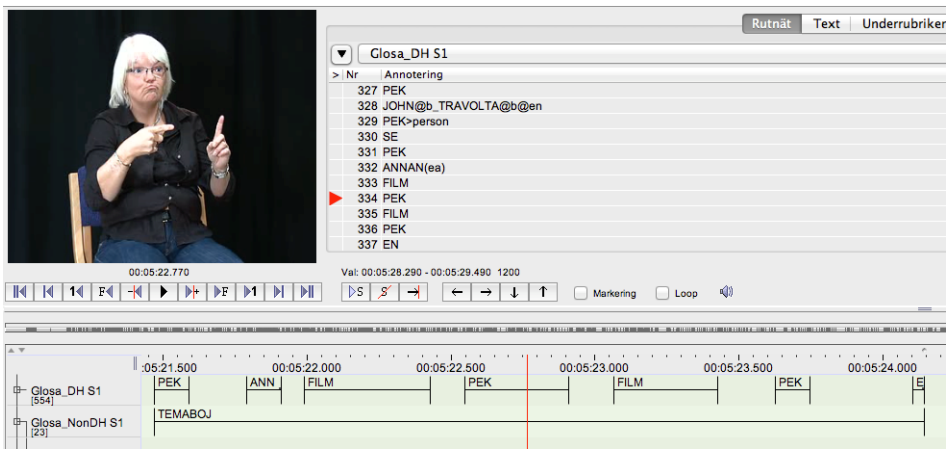
Figur 22 visar ett exempel på en pekboj (från eng. ”POINTER buoy”) (jfr. Liddell 2003; Nilsson, 2007) utförd med den icke-dominanta handen (här vänster) med en pekfingerhand som pekar framåt.[47] Kännetecknande för en pekboj är att den pekar mot ett element i texten. Pekbojar annoteras med PEKBOJ. Observera att pekbojen kan vara förvillande lik den pronominella pekningen PEK när den utförs med den icke-dominanta handen, som i Figur 32 nedan. Vid osäkert läge går det alltid ta till nödlösningen att lägga till ett @z efter.



Figur 22. Exempel på en pekboj, annoterad PEKBOJ i NonDH-raden

4.3 Temaboj

Figur 23 visar exempel på en temaboj (från eng. ”THEME buoy”) (jfr. Liddell 2003; Nilsson, 2007) utförd med den icke-dominanta handen (här vänster). Den består av en handform med ett (vanligtvis) uppåtriktat pekfinger. Temabojens närvaro lyfter fram det ämne som avhandlas eller ska avhandlas i texten.[48] Temaboj annoteras med glosan TEMABOJ.



Figur 23. Exempel på en temaboj, annoterad med TEMABOJ i NonDH-raden

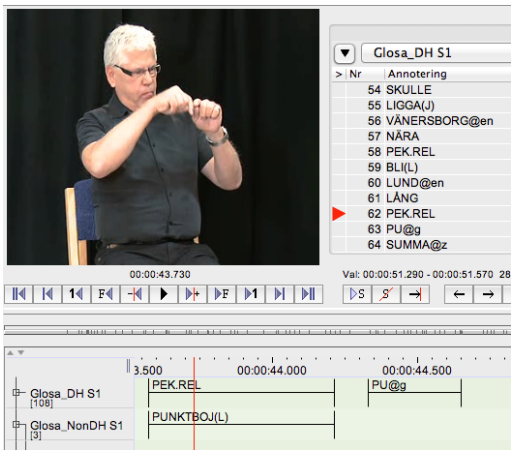
---

[47] *Bergman (2002) beskriver ett liknande tecken som TEMA-PEK i svenskt teckenspråk.*

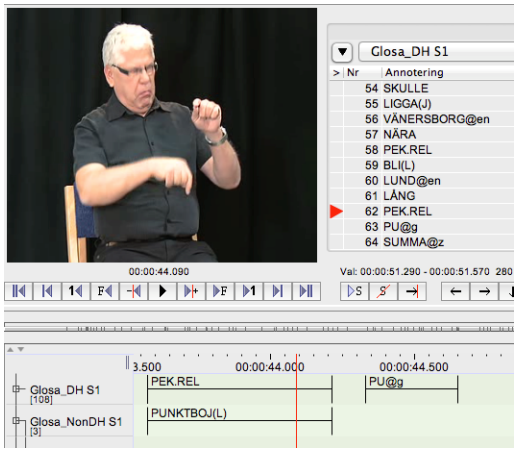
[48] *I sammanhanget i exemplet diskuterar den tecknande en skådespelare och hans medverkan i olika filmer och temabojen representerar skådespelaren.*

4.4 Punktboj

Figur 24 visar exempel på en punktboj (Vogt-Svendsen & Bergman, 2007) utförd med den icke-dominanta handen (här vänster) bestående av en handform med böjt pekfinger. Fingret kan också vara utsträckt, beroende på hur handen hålls i rummet. Punktbojen annoteras med glosan PUNKTBOJ följd av (L), dvs. PUNKTBOJ(L). Den kan också bestå av flat hand och annoteras då med tillägget (J), dvs. PUNKTBOJ(J). Från punktbojen gör dominanta handen, annoterad med PEK.REL[49], förflyttningar i rummet. Figur 24 och 25 visar tillsammans dominanta handens förflyttning neråt. Utförandet annoteras med PEK.REL. Dominanta handens förflyttning visar en relation mellan två punkter. I sammanhanget beskriver informanten sträckan mellan Göteborg, som representeras vid punktbojen i Figur 24, och Lund, som representeras vid förflyttningens slut i Figur 25.



Figur 24. Exempel på en punktboj, annoterad PUNKTBOJ(L) i NonDH-raden

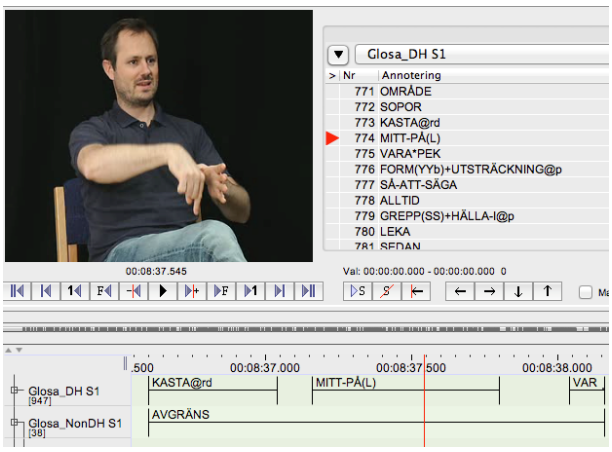


Figur 25. Visar slutet på dominanta handens förflyttning

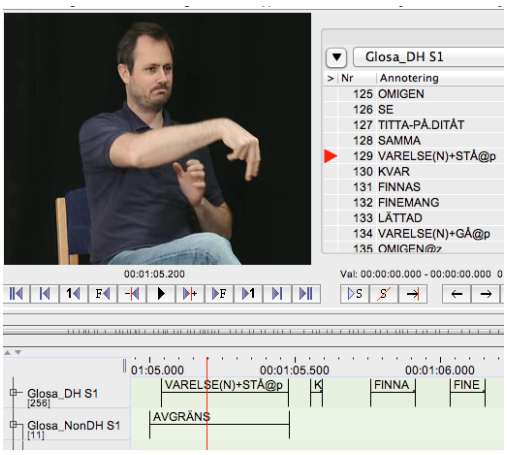
4.5 AVGRÄNS

Den icke-dominanta handen utför i Figur 26 ett exempel på det vi annoterar med glosan AVGRÄNS. Handformen utgörs vanligen av alla fingrar, som kan variera både med avseende på om de är samlade eller lätt åtskilda, och om de är böjda eller utsträckta. Kännetecknande för AVGRÄNS är att den ofta utförs när man berättar att man t.ex. bor i den inre delen av en by, eller att man kastar något mot ett avgränsat område (som i Figur 26, där den tecknande visar ett hål i mitten på ett bord). Den dominanta handen utför då tecken innanför den icke-dominanta handen. Dominanta handen kan också utföra tecken framför den icke-dominanta handen som i Figur 27.

[49] *REL är en förkortning av relation.*



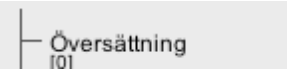
Figur 26. Exempel på AVGRÄNS, annoterad i NonDH-raden, med dominanta handen innanför icke-dominanta handen



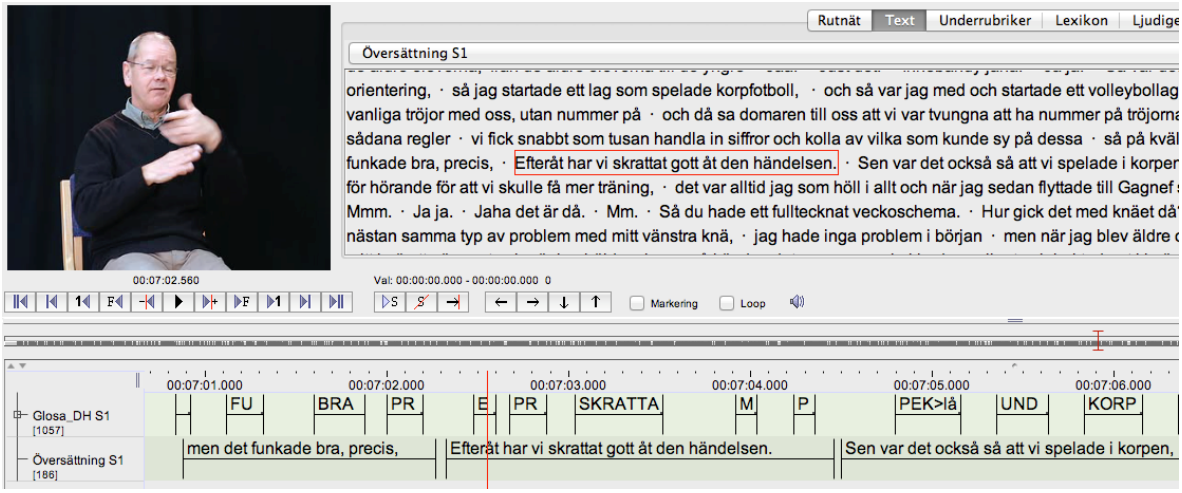
Figur 27. Exempel på AVGRÄNS, annoterad i NonDH-raden, med dominanta handen framför icke-dominanta handen

AVGRÄNS är en preliminär glosa, då alla förekomster ännu inte har analyserats. Anledningen till namnvalet är intrycket av att tecknet har ett samband med avgränsning mellan ”inre” och ”yttre” element, som ’innanför – utanför’.

5 Översättningsrad

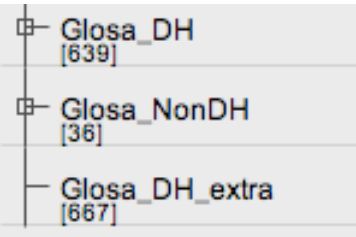


I översättningsraden annoteras en svensk översättning av texten. Man får då dela upp texten i celler i lämpliga längder, baserade på bl.a. fraser och satser eller meningar, och skriva översättningen som i Figur 28.

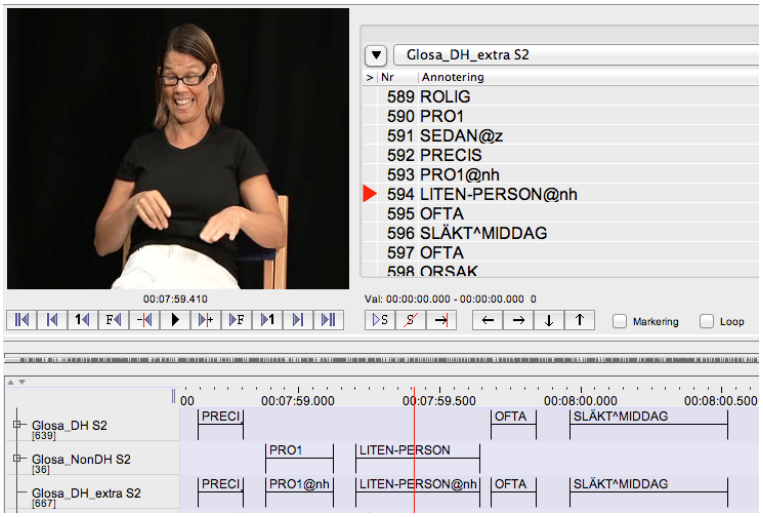


Figur 28. Exempel på översättning i Översättningsraden

6 Glosa_DH_extra



Raden Glosa_DH_extra är en kopierad rad från både Glosa_DH och Glosa_NonDH. I den raden hittar vi alltså annoteringar inte bara från Glosa_DH utan också från Glosa_NonDH, men med extra annoteringstillägg (därav ’extra’ i radnamnet). Ett exempel visas i Figur 29, tredje raden (som är den extra raden) där första glosan, PRECIS, och fjärde och femte glosan OFTA och SLÄKT^MIDDAG är kopior från första eller dominanta raden Glosa_DH. Andra glosan PRO1@nh och LITEN-PERSON@nh är kopior från andra eller icke-dominanta raden Glosa_NonDH men med tillägget @nh (förkortning av "non-dominant hand").

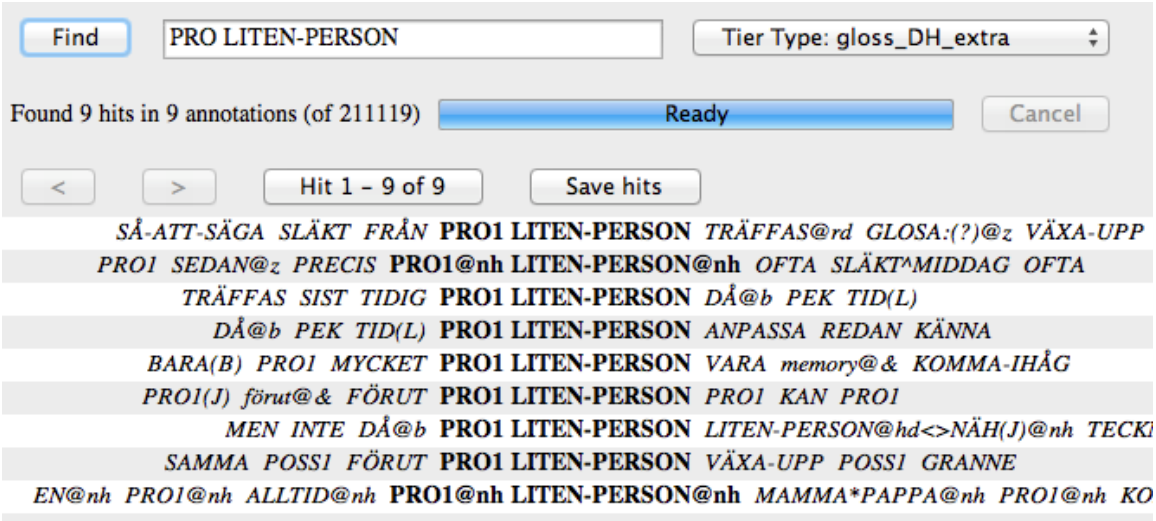


Figur 29. Annoteringssida med extra annotering av icke-dominanta tecken i GLOSA_DH_extra

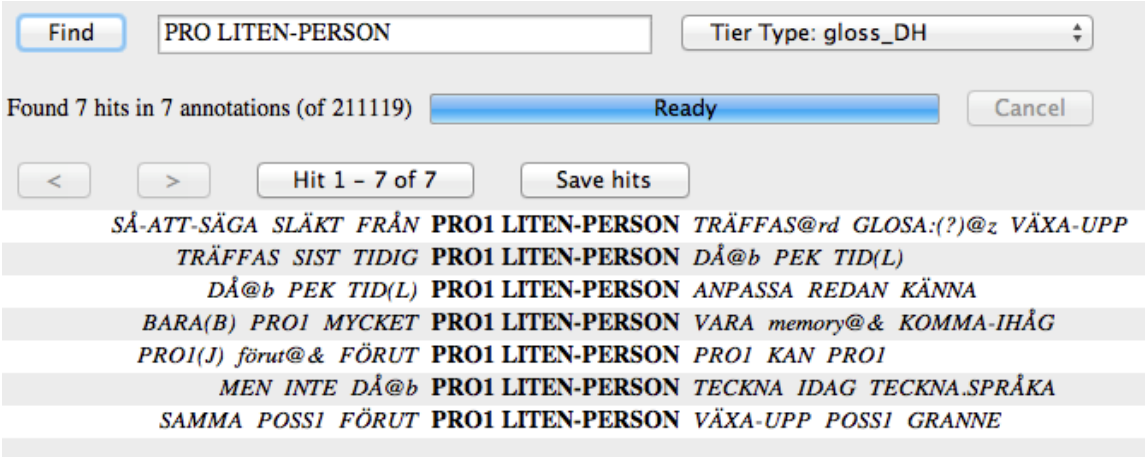
Avsikten med denna extra annotering är bland annat att dominanta och icke-dominanta tecken kommer med samtidigt vid studier av satsstrukturer eller hur ett bestämt tecken används i olika kontexter. Vid sökning i Glosa_DH kommer enbart dominanta tecken upp och vid sökning i Glosa_NonDH kommer enbart icke-dominanta tecken upp. Däremot kommer såväl dominanta som icke-dominanta tecken upp vid sökning i Glosa_DH_extra.

Figur 30 visar resultatet från en sökning i GLOSA_DH_extra, som genererat en konkordanssida med en uppsättning förekomster av teckensekvenser med den sökta sekvensen PRO1 LITEN-PERSON i fetstil. Antalet förekomster är nio, i två av dessa sekvenser utförs den sökta teckensekvensen av icke-dominanta tecken (markerade med @nh). Dessutom visas i andra raden hur icke-dominanta och dominanta handen

tillsammans bildar en sats från PRO1 till SLÄKT^MIDDAG. Denna rad kommer inte med på den andra konkordanssidan (Figur 31) efter sökning i Glosa_DH, som har endast sju förekomster istället för nio.

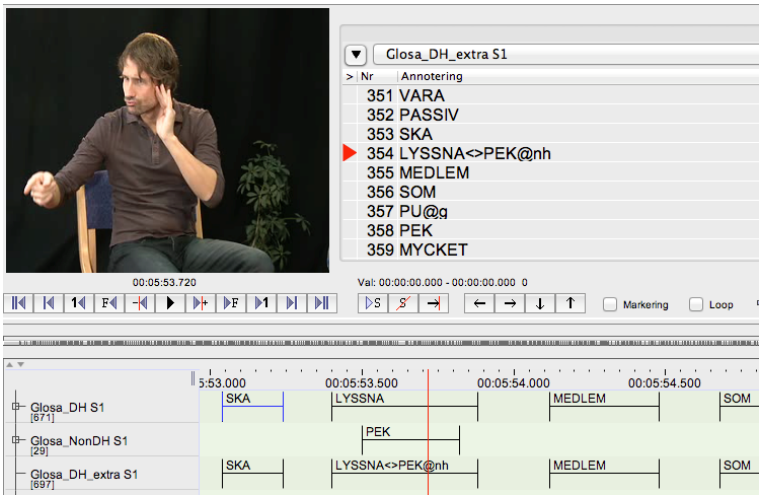


Figur 30. Konkordanssida med förekomster av icke-dominanta tecken

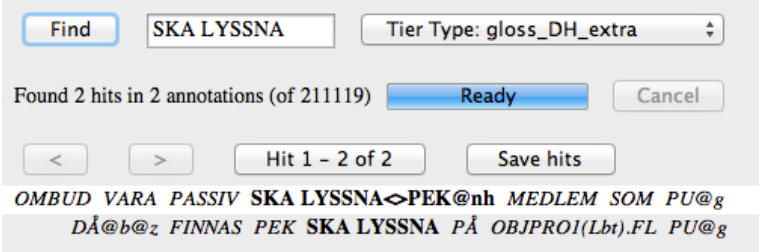


Figur 31. Konkordanssida utan förekomster av icke-dominanta tecken

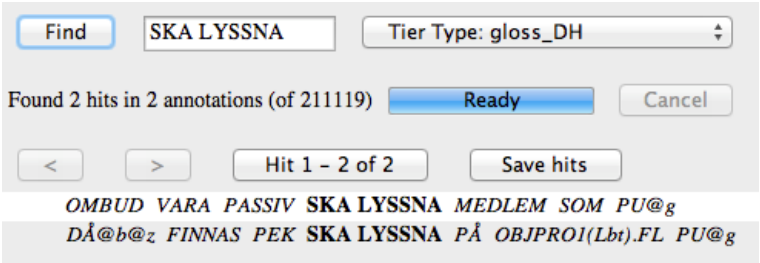
Ytterligare en avsikt med den extra annoteringen är överlappningar – när dominanta tecken och icke-dominanta tecken förekommer samtidigt (se figur 32). Överlappningen markeras med '<>' mellan dominanta tecknet LYSSNA och icke-dominanta tecknet PEK (med @nh). I Figur 33 med konkordanssidan från en sökning i Glosa_DH_extra kan man se överlappningen, men inte i Figur 34 från sökning i Glosa_DH.



Figur 32. Annoteringssida annoterad överlappning i Glosa_DH_extra

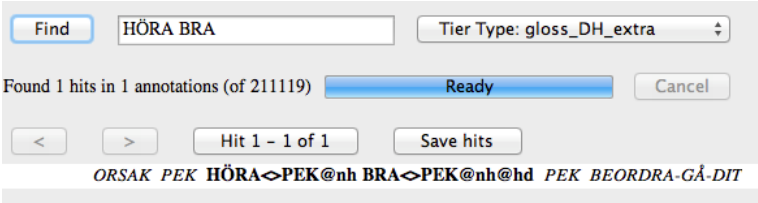


Figur 33. Konkordanssida med annoterad överlappning

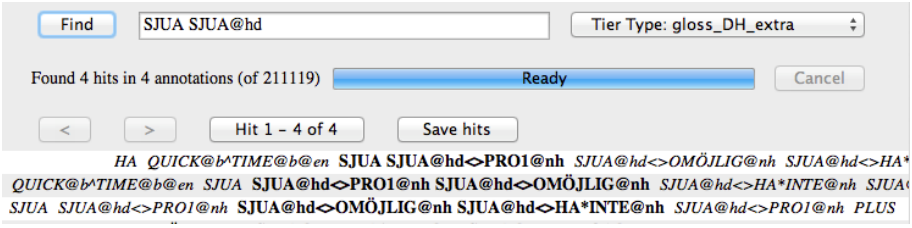


Figur 34. Konkordanssida utan annoterad överlappning

Det finns en tredje extra annotering som görs i Glosa_DH_extra. Det är när ena handen bibehålles i stationär position medan den andra handen producerar fler än ett tecken, som visas i Figur 35. Icke-dominanta PEK överlappar dominanta HÖRA och överlappningen bibehålles under dominanta BRA. Bibehållningen annoteras med @hd (förkortning av engelska 'hold'). Den dominanta handen kan även hållas i stationär position medan den icke-dominanta handen producerar tecken som visas i Figur 36. Andra raden visar att dominanta tecknet SJUA är bibehållet i stationär position medan den icke-dominanta handen producerar PRO1 och OMÖJLIG.



Figur 35. Konkordanssida med icke-dominanta handen i stationär position, annoterad med @nh@hd



Figur 36. Konkordanssida med dominanta handen i stationär position, annoterad med @hd

7 Annotering av icke-manuella komponenter

Med undantag av avsnittet om munnen är den följande texten om annotering av icke-manuella komponenter identisk med Wallin, Mesch och Nilsson (2011) Konventionerna har inte bearbetats eller utvecklats i samma grad som konventionerna för manuella delen. Konventionerna för den icke-manuella delen är således fortsatt huvudsakligen preliminära.

7.1 Mun

Annoteringskonventionen för munrörelser är lånad från Crasborn et al (2008). Munrörelserna annoteras efter fem aktivitetstyper. Den första är *Mouthings* och annoteras med M-type. Det är munrörelser som är lånade från svenska ord, som i tecken för ’billig’ (ID-nr 3377), ’brist’ (ID-nr 1690) eller ’tom’ (ID-nr 453). Den andra är *Adverbial mouth gestures* och annoteras med A-type. Det är munrörelser som beskrivs som orala adverb och som uttrycker något om storlek, såsom stor eller liten, ansträngdhet, såsom lätthet eller ansträngande, eller vikt, såsom lätt eller tung. Den tredje är *Semantically empty mouth gestures* och annoteras med E-type. Det är munrörelser som beskrivs som lexikala tillsammans med manuella delen. Exempel på dem är ’pinsam’ (ID-nr 170), ’lyckas’ (ID-nr 1553) eller ’leda till’ (ID-nr 1818). Den fjärde är *Enacting mouthgestures (mouth 4 mouth)* och annoteras med 4-type. Det är munrörelser som imiterar aktiviteter som tugga, spotta, blåsa eller bita. Den femte är Whole-face och annoteras med W-type. Det är munrörelser som uttrycker förvåning med gapande mun eller besvikelse med snörpt mun som del av ett ansiktsuttryck.

Utöver dessa fem tillkommer ytterligare tre annoteringar. Den första är till munrörelser som används vid uppbackning och annoteras B-type (efter engelska *backchannel*). Den andra är munrörelser som ej går att bestämma typ och annoteras med ’obestämd’. Den tredje är munrörelser som är obefintliga eller neutrala och annoteras därav ’neutral’.

Det förekommer situationer när en munrörelse påbörjar ett tecken före det tillhörande tecknet. Då annoteras den munrörelsen med passande aktivitetstyp följd av ’.regr’ (en förkortning för regressiv), som så här: BRUKA.regr. Det förekommer också situationer där munrörelsen fortsätter över nästa tecken. Då annoteras denna munrörelse med passande aktivitetstyp följd av ’.progr’ (förkortning för progressiv), som så här: NÄSTA.progr.[50]

För att sammanfatta:

M-TYPE	lånade från svenska ord
A-TYPE	orala adverb
E-TYPE	lexikal munrörelse
4-TYPE	imiterande munrörelse
W-TYPE	uttryck
B-TYPE	uppbackning
obestämd	går ej bestämma typ av munrörelse
neutral	obefintlig munrörelse

7.2 Ögon

Annotering av vad ögonen gör är uppdelad i två rader: blickriktning och öppningsgrad. I raden för blickriktning noteras aktiviteter hos ögonen, dvs. åt vilket håll den tecknande riktar blicken. I raden för öppningsgrad noteras aktiviteter hos ögonlocken, såsom när man blinkar eller kisar.

7.2.1 Blickriktning

I en dialog är blicken normalt riktad mot samtalspartnern. Vad som är höger eller vänster och upp eller ner är avhängigt av var den tecknande har sin normala blickriktning mot samtalspartnern. Därför är det väsentligt att man själv har noterat den normala blickriktningen, som i Figur 37, och sedan utifrån denna annoterar när blickriktningen avviker från normalläget åt sidorna eller upp och ner, som i Figur 38–41.

[50] *Annoteringskonventionen med regressiv och progressiv är lånad från Johnston (2015).*

Figur 37. Blicken är riktad mot adressaten

Figur 38. Blicken är riktad till höger om adressaten

Figur 39. Blicken är riktad till vänster om adressaten

Figur 40. Blicken är riktad uppåt

Figur 41. Blicken är riktad neråt

Figur 42. Blicken riktad höger och neråt

När den tecknande har blicken riktad åt t.ex. höger och ner, som i Figur 42, annoteras riktningen med kombinationen hö-ner.

Sammanfattningsvis annoteras blickriktningen enligt följande:

adr	blicken är riktad mot adressaten
hö	blicken är riktad till höger om adressaten
vä	blicken är riktad till vänster om adressaten
ner	blicken är riktad neråt
upp	blicken är riktad uppåt

7.2.2 Öppningsgrad

Annoteringen av ögonens öppningsgrad utgår från hur den tecknande har sina ögonlock i normalläge. Detta varierar mellan olika personer. En del har smala ögon och andra har mer öppna ögon. Därför ska man innan annoteringen försöka lägga märke till ögonlockens normalläge, som i Figur 43. Utifrån detta annoterar man när ögonlocken ändras från normalläget, som att de höjs så att ögonen vidgas (Figur 44) eller sänks så de kisar (Figur 45).

Det finns ytterligare två tillfällen där öppningsgraden annoteras. De är blink (Figur 46) och sluten (Figur 47). Blink väljs när ögonen är slutna upp till fyra eller fem bildrutor. Är ögonen slutna längre än fem bildrutor annoteras det som slutna, vilket kan vara en långsam blinkning eller att den tecknande blundar.

Ögonlocken i normalläge annoteras normalt inte.

Figur 43. Den tecknandes ögonlock i normalläge

Figur 44. Den tecknande har vidgade ögon

Figur 45. Den tecknande har kisande ögon[51]

[51] *Kisning kombineras vanligen med kinderna rör sig uppåt (se Figur 44).*

Figur 46. Den tecknande blinkar

Figur 47. Den tecknande har slutna ögon

Sammanfattningsvis annoteras öppningsgraden enligt följande:

bl	blink
sl	slutna
vid	vidgade
kis	kisande

7.3 Ögonbryn

I ögonbrynsraden noteras aktiviteter hos ögonbrynen. Inför annoteringen ska man försöka lägga märke till var den tecknande har sina ögonbryn i normalläge, som i Figur 48. Utifrån normalläget annoterar man när ögonbrynen ändras från normalläget, som i Figur 49 och 50, om de höjs eller sänks. Det är dessa två händelser man noterar i ögonbrynsraden. Ibland kan det vara svårt att avgöra om en förändring har skett. Ett tips är att se om det har bildats rynkor i pannan eller vid näsroten.

Figur 48. Den tecknandes ögonbryn i normalläge

Figur 49. Ögonbrynen i höjt läge

Figur 50. Ögonbrynen i sänkt läge

Ögonbrynen i normalläge annoteras inte.

Ögonbrynen annoteras alltså enligt följande:

höjda	ögonbrynen i höjt läge
sänkta	ögonbrynen i sänkt läge

7.4 Näsa / Kinder

I raden för kind och näsa kan man notera aktiviteter hos kinderna och näsan.

Figur 51. Kinderna höjs

Figur 52. Den tecknande rynkar på näsan

Kinderna annoteras med:

upp kinderna höjs

Näsan annoteras med:

rynk rynkar på näsan

7.5 Huvud

I huvudraden noteras huvudets aktiviteter, dessa beskrivs i två plan. Huvudet gör dels vridrörelser åt sidorna och dels böjningsrörelser framåt och bakåt, eller åt sidorna. Detta är förmodligen den jämförelsevis mest komplicerade icke-manuella delen att annotera. Det är ofta tätt mellan olika aktiviteter och det blir en avvägningsfråga när någon aktivitet ska noteras. Det händer att den tecknande gör små huvudrörelser utan att allt nödvändigtvis ska noteras. Riktlinjen kan då sägas vara att en vridning eller en böjning bör gå över minst tio bildrutor för att noteras.

Utgångspunkten för notering av huvudets aktiviteter är hur den tecknande normalt har sitt ansikte (eller näsan) riktat mot samtalspartnern som i Figur 53. Utifrån detta noterar man sedan de olika aktiviteterna som i Figur 54–58.

Figur 53. De tecknandes normalposition med ansiktet vänt mot samtalspartnern

Vridning är en huvudrörelse i horisontalled, dvs. vridning åt vänster eller åt höger, som i Figur 53. Upprepas vridrörelserna kontinuerligt (som när man negerar med huvudet) noteras aktiviteten som skakning.

Figur 54. Huvudet vrids åt höger

Böjning är en huvudrörelse i vertikalled, som att böja huvudet framåt eller bakåt eller åt sidorna som i Figur 54–55. Dessutom har vi nick (en kort huvudböjning fram och åter).

Figur 55. Huvudet böjs framåt

Figur 56. Huvudet böjs bakåt

Det finns ytterligare två huvudrörelser att notera. Det är när huvudet förs framåt så att hakan närmast sticker ut (Figur 56) eller förs bakåt så att hakan förs in mot halsen (Figur 57).

Figur 57. Huvudet förs framåt

Figur 58. Huvudet förs bakåt

Ibland förekommer det att huvudet böjs åt höger och vrids åt vänster samtidigt. Då får man kombinera noteringar med t.ex. böj-hö_vrid-vä.

Huvudrörelser noteras sammanfattningsvis enligt följande:

vrid.vä	huvudet vrids åt vänster
vrid.hö	huvudet vrids åt höger
böj.fram	huvudet böjs framåt
böj.bak	huvudet böjs bakåt
böj.vä	huvudet böjs åt vänster
böj.hö	huvudet böjs åt höger
fram	huvudet förs framåt
bak	huvudet förs bakåt
skak	skakar på huvudet eller gör upprepade vridrörelser
nick	nickar eller huvudet böjs kort fram
nick.bak	huvudet böjs kort bak
nick.uppr	nickar upprepade gånger

7.6 Kropp

I kroppsraden noteras aktiviteter hos kroppen. Med kroppen avses här den del av kroppen som är från axlarna och ner till höften. Liksom med huvudet kan aktiviteterna beskrivas i två plan. Kroppen gör dels vridningar åt sidorna och dels lutningar framåt och bakåt eller åt sidorna (jfr. huvudrörelserna). Dessutom gör kroppen eller axlarna en rörelse uppåt. Det blir här också en avvägningsfråga när kroppsrörelserna ska noteras. Det händer att den tecknande gör små kroppsrörelser utan att allt nödvändigtvis ska noteras. Riktlinjen kan sägas vara att kroppsförändringen bör vara bibehållen över flera bildrutor för att noteras. Man annoterar när kroppsförändringen har någon grammatisk funktion.

Det vanliga utgångsläget eller riktlinjen för noteringar av kroppens olika rörelser är att den tecknande sitter relativt upprätt, med hela framsidan vänd mot samtalspartnern, som i Figur 59. Men andra kroppspositioner kan förekomma och då får man ta ställning till vad som är normalposition och utifrån den annotera när kroppen ändras från normalpositionen.

Figur 59. Den tecknandes kropp i normalposition.

Figur 60–62 visar några exempel på kroppsrörelser som noteras utifrån normalpositionen i Figur 59.

Figur 60. Kroppen böjs bakåt

Figur 61. Kroppen vrids åt höger

Figur 62. Kroppen böjs åt höger[52]

[52] *Den tecknandes normala kroppsposition är ungefär som den i Figur 59.*

Kroppsrörelser noteras som följande:

vrid.vä	kroppen vrids åt vänster
vrid.hö	kroppen vrids åt höger
fram	kroppen böjs framåt
bak	kroppen böjs bakåt
vä	kroppen böjs åt vänster
hö	kroppen böjs åt höger
axel.upp	axlarna rör sig uppåt
axel.ryck	rycker på axlarna eller axlarna rör sig kort uppåt

8 Referenser

Bergman, B. (1982). ”Transkription av texter” i Teckenspråkstranskription. *Forskning om Teckenspråk X*. Institutionen för lingvistik, Stockholms universitet.

Bergman, B. (1990). Föreläsningsanteckningar om verb och adjektiv: Några morfologiska processer i det svenska teckenspråket. Institutionen för lingvistik, Stockholms universitet

Bergman, B. (2002). Föreläsningsanteckningar om pekningar. Institutionen för lingvistik, Stockholms universitet.

Bergman, B. (2012). Barns tidiga teckenspråksutveckling. *Forskning om Teckenspråk XXII*. Institutionen för lingvistik, Stockholms universitet.

Bergman, B. & Dahl, Ö. (1994). Ideophones in Sign Language? The place of reduplication in the tense-aspect system of Swedish Sign Language. In: Bache, C., Basbøll, H. & Lindberg, C.E. (eds.). *Tense, Aspect and Action. Empirical and Theoretical Contributions to Language Typology*, pp. 397-422. New York, NY: Mouton de Gruyter

Bergman, B. & Wallin, L. (2001). A preliminary analysis of visual mouth segments in Swedish Sign Language. In Boyes Braem, P. & Sutton-Spence, R. (eds) (2001): *The Hands are the Head of the Mouth. The Mouth as Articulator in Sign Languages.* (International studies on sign language and communication of the deaf; Vol 39.) Hamburg: Signum-Verlag.

Boyes Braem, P. & Sutton-Spence, R. (eds) (2001). *The Hands are the Head of the Mouth. The Mouth as Articulator in Sign Languages.* (International studies on sign language and communication of the deaf; Vol 39.) Hamburg: Signum-Verlag.

Börstell, C. (2011). *Revisiting Reduplication. Toward a description of reduplication in predicative signs in Swedish Sign Language*. Institutionen för lingvistik, Stockholms universitet. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-63510 (2 september 2014)

Crasborn, O., E. van der Kooij, D. Waters, B. Woll & J. Mesch. (2008). Frequency distribution and spreading behavior of different types of mouth actions in three sign languages. *Sign Language & Linguistics 11:1*, pp. 45-67.

Hedberg, T. (1989). Persontecken. Deras härkomst, bildningssätt och användning. *Forskning om Teckenspråk X*. Institutionen för lingvistik, Stockholms universitet.

Johnston, T. (2015). *Auslan corpus annotation guidelines.* Unpublished manuscript. Centre for Language Sciences, Department of Linguistics, Macquarie University, Sydney, Australia.

Liddell, S. (2003). *Grammar, Gesture, and Meaning in American Sign Language*. Cambridge: Cambridge University Press.

Liddell, S. K., Vogt-Svendsen, M. & Bergman, B. (2007). “A crosslinguistic comparison of buoys. Evidence from American, Norwegian, and Swedish Sign Language.” In Vermeerbergen, M., Leeson, L. & Crasborn, O. (eds.), *Simultaneity in Signed Languages: Form and Function*. Amsterdam/Philadelphia: John Benjamins Publishing Company. pp. 187–215.

MacWhinney, B. (2012). URL: http://childes.psy.cmu.edu/manuals/CHAT.pdf

Mesch, J. (2011). *Att använda ELAN. Bruksanvisning för annotering och studie av teckenspråkstexter*. Version 3. Institutionen för lingvistik, Stockholms universitet. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-64289 (2 september 2014)

Mesch, J., Rohdell, M. & Wallin, L. (2012). *Annoterade filer för svensk teckenspråkskorpus. Version 1*. Avdelningen för teckenspråk, Institutionen för lingvistik, Stockholms universitet. URL: http://www.ling.su.se/teckenspraksskorpus (2 september 2014)

Mesch, J., Rohdell, M. & Wallin, L. (2014). *Annoterade filer för svensk teckenspråkskorpus. Version 2*. Avdelningen för teckenspråk, Institutionen för lingvistik, Stockholms universitet. URL: http://www.ling.su.se/teckenspraksskorpus (2 september 2014)

Mesch, J., Wallin, L., Nilsson, A-L. & Bergman, B. (2012). *Datamängd. Projektet Korpus för det svenska teckenspråket 2009-2011 (version 1)*. Avdelningen för teckenspråk, Institutionen för lingvistik, Stockholms universitet. URL: http://www.ling.su.se/teckenspraksskorpus (2 september 2014)

Metzger, M. (1995). Constructed dialogue and constructed action in American Sign Language. In C. Lucas (ed), *Sociolinguistics in deaf communicaties.* Washington, D.C., Gallaudet University Press.

Nilsson, A-L. (2007). “The non-dominant hand in a Swedish Language discourse.” In Vermeerbergen, M., Leeson, L. & Crasborn, O. (eds): *Simultaneity in Signed Languages: Form and Function*. Amsterdam/Philadelphia:John Benjamins Publishing Company. pp. 163–185.

Nonhebel, A., Crasborn, O. & van der Kooij, E. (2004). *Sign language transcription conventions for the ECHO project.* URL: http://www.academia.edu/1046424/Sign_language_transcription_conventions_for_the_ECHO_project (15 november 2012)

Roush, D. (2007). ”Indirectness strategies in American Sign Language Requests and Refusals: Deconstructing the Deaf-as-Direct Stereotype.” In Metzger, M. & Fleetwood, E. (eds.): *Translation, Sociolinguistic, and Consumer Issues in Interpreting*. Washington, D.C., Gallaudet University Press. pp. 103–156.

Simper-Allen, P. (2013). *Avbildande verbkonstruktioner i svenskt teckenspråk.: Handformskategorier inom "Cut and Break"-domänen*. Institutionen för lingvistik, Stockholms universitet. URL: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-98624 (2 september 2014)

Svenskt teckenspråkslexikon. (2008--).
URL: http://www.ling.su.se/teckenspraksleksikon (2 september 2014)

Vogt-Svendsen, M. & Bergman, B. (2007). ”Point buoys. The weak hand as a point of reference for time and space.” In Vermeerbergen, M., Leeson, L. & Crasborn, O. (eds.), *Simultaneity in Signed Languages: Form and Function*. Amsterdam/Philadelphia: John Benjamins Publishing Company. pp. 217–235.

Wallin, L. (1994). *Polysyntetiska tecken i svenska teckenspråket*. Institutionen för lingvistik, Stockholms universitet.
URL: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-20016 (2 september 2014)

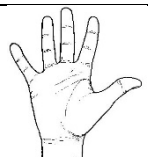
Wallin, L., Mesch, J. & Nilsson, A-L. (2011). *Transkriptionskonventioner för svenska teckenspråket. Version 3.* Institutionen för lingvistik, Stockholms universitet.
URL: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-69190 (2 september 2014)

Bilaga 1. Handformssymboler

A	
Ao	
B	
D	
G	
H	
I	
It	
J	

Jb	s
Jt	
Jv	
L	
Lb	
Lbs	
Lbt	
Lt	tummen framböjd som Jt
Lv	

Lvs	
Lvt	
N	
O	
Q	
R	
S	
V	
Vb	
Vbs	

Vbt	
Vt	
W	
Y	
Yb	
Ybt	
Yt	
Z	
4	

Bilaga 2. Klassifikatoriska tecken

ANSIKTE@kl

AVLÅNG-LITEN@kl

AVLÅNG@kl

BRED@kl

FYRKANT@kl

FÄLT@kl

GLOB@kl

GRUPP@kl

HUS@kl

HÖJD@kl

KLÄDD@kl

KOMPENDIUM@kl

LITEN-PERSON@kl

OMRÅDE@kl

PERIOD@kl

PERSON@kl

PERSON.FL@kl

PERSON.MULTI@kl

RUND@kl

Bilaga 3. Polysyntetiska tecken

Entitet

ENTITET(G)	ENTITET(L)	ENTITET(N)	ENTITET(Vbt)
ENTITET(It)	ENTITET(Lb)	ENTITET(O)	ENTITET(Y)
ENTITET(J)	ENTITET(Lbt)	ENTITET(S)	ENTITET(Yb)
ENTITET(Jb)	ENTITET(Lt)	ENTITET(V)	ENTITET(Ybt)
ENTITET(Jv)	ENTITET(Lvt)	ENTITET(Vbs)	ENTITET(4)

Varelse

VARELSE(J)	VARELSE(N)	VARELSE(Y)
VARELSE(L)	VARELSE(V)	VARELSE(Yb)
	VARELSE(Vb)	VARELSE(4)

Hand

HAND(A)	HAND(Jb)	HAND(Q)	HAND(Y)
HAND(Ao)	HAND(Jt)	HAND(Qo)	HAND(Yb)
HAND(G)	HAND(Jv)	HAND(S)	HAND(Ybt)
HAND(H)	HAND(Lvt)	HAND(Vbt)	
HAND(J)	HAND(Lvs)		

Grepp

GREPP(G)

GREPP(Q)

Form

FORM(Ao)	FORM(Hr)	FORM(L)	FORM(S)
FORM(B)	FORM(Ho)	FORM(Lb)	FORM(Vbt)
FORM(D)	FORM(J)	FORM(Lbt)	FORM(Y)
FORM(G)	FORM(Jb)	FORM(Lvs)	FORM(Yb)
FORM(H)	FORM(Jv)	FORM(O)	FORM(Ybt)

Övriga

KOLLEKTIV(YYb)

VÄTSKA(Y)

I serien **Forskning om teckenspråk** (FOT) har följande nummer utkommit. Samtliga rapporter är fritt tillgängliga på avdelningens hemsida: www.ling.su.se/teckensprak

FOT I, 1976
Ahlgren, Inger. Rapport om planering och förarbete i projektet "Tidig språklig kognitiv utveckling hos döva och gravt hörselskadade".

FOT II, 1978
Bergman, Brita. The first national symposium on sign language research and teaching, Chicago May 30 – June 4, 1977.
Ulfsparre, Sten. Teaching sign language to hearing parents of deaf children.
Ahlgren, Inger. Early linguistic cognitive development in the deaf and severely hard of hearing.

FOT III, 1979
Bergman, Brita. Dövas teckenspråk – en inledning.

FOT IV, 1978
Bergman, Brita. Current developments in sign language research in Sweden.

FOT V, 1979
Ahlgren, Inger. Om språkutveckling hos hörande och döva.

FOT VI, 1980
Ahlgren, Inger. Projektet "Tidig språklig kognitiv utveckling hos döva och gravt hörselskadade": Arbetsätt och erfarenheter.
Ahlgren, Inger. Döva barn och vuxna döva.

FOT VII, 1980
Ahlgren, Inger. Döva barns teckenspråk.

FOT VIII, 1982
Wallin, Lars. Sammansatta tecken i svenska teckenspråket.

FOT IX, 1982.
Ozolins, Birgitta. Lördagsskolan 1977–78.

FOT X, 1982
Bergman, Brita. Teckenspråkstranskription.

FOT XI, 1982
Bergman, Brita. Sign typology.
Bergman, Brita. Verbs and adjectives: Some morphological processes in Swedish Sign Language.
Bergman, Brita. On localisation in the Swedish Sign Language.
Bergman, Brita. Några satstyper i det svenska teckenspråket.

FOT XI:2, 1982
Bergman, Brita. Några satstyper i det svenska teckenspråket.

FOT XII, 1984
Svartholm, Kristina. Döva och samhällets skrivna språk. En forskningsöversikt och en tillbakablick.

FOT XIII, 1984
Ahlgren, Inger. Döva barn och skriven svenska.

FOT XIV, 1984
Bergman, Brita. En jämförelse mellan några dynamiska och statiska satser i svenska och teckenspråk.
Bergman, Brita & Wallin, Lars. Satsstruktur i det svenska teckenspråket.
Bergman, Brita. Verb och adjektiv: Några morfologiska processer i svenska teckenspråket.
Ahlgren, Inger. Persondeixis i svenska och i teckenspråk.

FOT XV, 1987
Thyrén, Louise. Döva i Sverige – en delkultur?

FOT XVI, 1989
Hedberg, Tomas. Persontecken. Deras härkomst, bildningssätt och användning.

FOT XVII, 1992.
Bergman, Brita. Teckenspråket – ett svenskt minoritetsspråk. Föreläsning vid Stockholms universitets installations- och promotionshögtidighet den 25 september 1992.

FOT XVIII, 1994
Mesch, Johanna. Dövblindas teckenspråk. En pilotstudie.

FOT XIX, 1995
Bergman, Brita. Manuell och ickemanuell negation i svenska teckenspråket.

FOT XX, 1996
Ahlgren, Inger. Om dövblindfödda barns möjlighet till språkutveckling.

FOT XXI, 1998
Ahlström, Margareta & Svartholm, Kristina. Barndomshörselskadades erfarenheter och upplevelser av tvåspråkighet. En pilotstudie.

För senare FOT-nummer, se omslagets baksida.

Rapportserie Forskning om teckenspråk (FOT)
Redaktion: Brita Bergman, Johanna Mesch

FOT XXII, 2012	Bergman, Brita. Barns tidiga teckenspråksutveckling
FOT XXIII, 2013	Simper-Allen, Pia. Avbildande verbkonstruktioner i svenskt teckenspråk
FOT XXIV, 2015	Mesch, Johanna. Svensk teckenspråkskorpus - dess tillkomst och uppbyggnad Wallin, Lars & Mesch, Johanna. Annoteringskonventioner för teckenspråkstexter

ISSN: 99-0256266-3

Institutionen för lingvistik
Avdelningen för teckenspråk
Stockholms universitet 106 91 Stockholm www.ling.su.se/teckensprak

Stockholms universitet