

Institutionen för lingvistik vid Stockholms universitet deltar i Swe-Clarin genom två av sina avdelningar: datorlingvistik och teckenspråk.

Avdelningen för datorlingvistik bedriver främst forskning kopplad till förstaspråksinlärning och länkning av parallellkorpusar (mer information finns på <http://www.ling.su.se/english/research/research-areas/research-in-computational-linguistics>). Avdelningen tillgängliggör ett flertal korpusar och verktyg av potentiellt intresse för Swe-Clarin, inklusive Stagger (en satsdelstaggare med modeller för svenska och isländska), SIC (Stockholm Internet Corpus) och SUC-CORE (en delmängd av SUC med koreferensannotering för nominalfraser). En tidigare infrastrukturell satsning var the Stockholm-Umeå Corpus (SUC) som omfattar en miljon ord med manuellt kontrollerade ordklasstaggar, och vars första version utvecklades 1989–97.

Distributionen av SUC övergick till Språkbanken 2008 men korpusen underhålls fortfarande av institutionen; den senaste versionen (SUC 3.0) släpptes 2012. För mer information om resurser, se <http://www.ling.su.se/english/nlp>.



Mycket av dagens forskning fokuserar på lexikografi och korpusar för teckenspråk, framför allt svenskt teckenspråk (SSL). Ett genombrott i representation av teckenspråksdata skedde i mitten av 1990-talet när det blev möjligt att spela video på persondatorer, vilket ersatte videoband. Dessutom har de system som utvecklats för detaljerad och sökbar annotering av teckenspråksvideo, som till exempel ELAN (<https://tla.mpi.nl/tools/tla-tools/elan/>), revolutionerat hur teckenspråkskorpusar kan representeras och analyseras. Arbetet med SSL-lexikonet inleddes 1988 och det omfattar nu cirka 15 000 tecken. SSL-korpusen påbörjades 2003 och innehåller för närvarande omkring 24 timmar med annoterad video från 42 teckenspråksanvändare av delvis spontana dialoger och monologer med berättelser och eliciteringar. Annoteringarna inkluderar förklaringar av tecken och översättningar till svenska på yttrandnivå (se bild). Andra resurser inkluderar en korpus för taktill teckenspråk för personer med dövblindhet, och en inlärarkorpus för SSL. För mer information om resurserna, se <http://www.ling.su.se/teckenspraksresurser>.

En sommar som lider mot sitt slut

Större delen av den nationella samordningen är åter på plats: Caspar har återvänt från sin båttripp på Donau, Stefan är tillbaka från sin båttripp på Disko Bay och Lars har fjordsicksackat sig hem från norra Skandinavien. Nina är däremot ledig till slutet av september. Sedan senaste nyhetsbrevet har Språkbanken hunnit fylla 40 år (vilket firades med såväl informativa som välsmakande festligheter), Swe-Clarin har haft ett virtuellt möte och Storbritannien ingår numera i CLARIN som observatör.

Hösten utlovar spännande CLARIN-aktiviteter såväl nationellt som internationellt (se kalendariet nedan). Vi i samordningen ser fram emot att träffa många av er snart!



Vintern glider närmare ...

Kalender

- 5–6 oktober: Nordic Clarin Network-workshop i anslutning till Språkbankens höstworkshop med tema historiska resurser.
- 7–8 oktober: den nationella samordningen besöker Swe-Clarin-centrum i Stockholm och Uppsala.
- 9 oktober: workshop om digitalisering anordnad av Digisam och KB.
- 15–17 oktober: CLARIN:s årliga konferens i Wrocław.
- 11 november: SND:s höstworkshop på temat "Nya villkor för forskning". Alla är välkomna.
- 16–17 november: Workshop och möte mellan Swe-Clarin:s parter i Stockholm.
- 19–20 november: möte för CLARIN ERIC:s General Assembly i Köpenhamn.
- 11 december: virtuellt möte för Swe-Clarin-parterna kl 10–12.

Parter

Swe-Clarin har nio parter från Lund, Göteborg, Linköping, Stockholm och Uppsala, på såväl universitet som myndigheter. En lista och beskrivning av alla parter hittar ni här:

<http://sweclarin.se/swe/centrum>

Nyheter

Vi kommer inte att fortsätta spamma. Om ni vill få vidare information om Swe-Clarin anmäl er till e-postlistan här: http://lists.sweclarin.se/mailman/listinfo/news_lists_sweclarin.se

