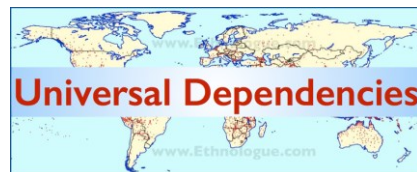


Datorlingvistikgruppen vid Uppsala universitet har en stark empirisk inriktning med särskilt fokus på flerspråkiga system, särskilt maskinöversättning, och system för grammatisk analys av text, närmare bestämt dependensbaserad parsning. Ett annat fokusområde är digitala humaniora med projekt om igenkänning av handskriven text, bearbetning av historiska texter och historiska chiffer. Gruppen har varit med och utvecklat ett antal verktyg och resurser, bland annat MaltParser (datadriven dependensparser), UPlug (verktygslåda för länkning av parallellkorpusar), Svensk trädbank (syntaktiskt uppmärkt korpus) och OPUS (flerspråkig parallellkorpus). Här nedanför beskriver vi några pågående projekt som är relevanta för CLARIN.



Inom projektet Universal Dependencies jobbar intresserade från olika håll med att utveckla likformig trädbanksuppmärkning för många olika språk. Källkoden är öppet tillgänglig. Detta kommer att underlätta spridningen av parsningsteknologi som kan stödja forskning inom humaniora och samhällsvetenskap på flera olika språk. Hittills har projektet släppt trädbanker för 18 språk (inklusive svenska).



Inom Gender & Work-projektet forskar historiker på vad män och kvinnor försörjde sig på i det tidigmoderna Sverige (1550-1800). För närvarande exciperas informationen genom att för hand leta igenom stora textmängder efter stycken som beskriver arbete. Vi utvecklar tekniker som ska underlätta denna process genom att automatiskt leta fram fraser som med stor sannolikhet beskriver arbete. Teknikerna bygger på stavningsnormalisering och lingvistisk analys.



Projektgruppen From Quill to Bytes arbetar med att analysera historiska handskrivna dokument med hjälp av metoder från bildanalys och språkteknologi. Målet är att utveckla metoder för att hitta språkliga enheter direkt i bilder på manuskript, för automatisk transkription av manuskript och för identifiering av skribenten, stilen eller åldern hos manuskript.



Tusentals chifferade historiska manuskript ligger glömda i bibliotek och arkiv. Vi utvecklar datorbaserade verktyg för automatisk dechiffriering av historiska chiffer. Projektet omfattar systematisk igenkänning av olika chifertyper, utveckling av algoritmer för dechiffriering och skapandet av språkmodeller och mönsterlexikon för tidiga varieteter av europeiska språk.

Besök på CiltLab i Linköping

I maj fortsatte den nationella samordningens besök på Swe-Clarins centrum som föreslogs på avsparken. Besöken syftar till att erbjuda centrumen och den nationella samordningen en möjlighet att lära känna varandra närmare och att kunna diskutera hur Swe-Clarins kan fungera som effektivast. Fredagen den 29:e reste Lars, Caspar och Stefan till Linköping för ett trevligt möte med centrumföreståndare Magnus Merkel och expert Lars Ahrenberg.

Swe-Clarins-centrumet vid Linköpings universitet ligger vid CiltLab (Cognition, Interaction and Language Technology) på avdelningen för Human-Centred Systems, Institutionen för datavetenskap. Där fick vi en presentation av den forskning och undervisning som bedrivs inom språkteknologi av de nio språkteknologer som är verksamma vid CiltLab, och en genomgång av de nio resurser som centrumet kommer att göra tillgängligt för CLARIN. De bredare diskussionerna kring Swe-Clarins rörde bland annat skillnaden mellan K-centrum och L-centrum inom CLARIN, och möjligheten att ta fram en basuppsättning med verktyg för tokenisering, lemmatisering, taggning och så vidare. Den senare frågan kommer att tas upp vidare på det virtuella mötet 12 juni.

Besök vid Swe-Clarins övriga centrum kommer förhoppningsvis att äga rum under andra halvan av september.

På det europeiska planet är det värt att notera att svenska forskare nu också har tillgång till CLARIN-tjänster med federerad inloggning (dvs. det går att logga in med sitt konto från hemlärosätet). Preliminära tester tyder på att uppkopplingen via eduGAIN fungerar bra. En annan nyhet är att Leif-Jöran Olsson från Språkbanken har tagit över ansvaret för att utveckla CLARIN:s innehållssök.

I och med detta tar nyhetsbrevet sommaruppehåll och återkommer i augusti. Trevlig sommar!

Kalender

- 9-10 juni: Nordic Clarin Network-workshop.
- 12 juni: Virtuellt möte för Swe-Clarins-parterna kl 10-12.
- 5-6 oktober: Nordic Clarin Network-workshop i anslutning till Språkbankens höstworkshop med tema historiska resurser.
- 11 november: SND:s höstworkshop på temat "Nya villkor för forskning". Alla är välkomna.
- 19-20 november: Swe-Clarins-stormöte i Stockholm, Swe-Clarins-workshop i anslutning på fredag eftermiddag. Alla är välkomna.

Partners

Swe-Clarins har nio parter från Lund, Göteborg, Linköping, Stockholm och Uppsala, på såväl universitet och 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-Clarins anmäl er till e-postlistan här: http://lists.sweclarin.se/mailman/listinfo/news_lists_sweclarin.se

