

Digitaliserad 1700-talstext som lexikaliskt verktyg – utmaningar och möjligheter

Sofie Johansson & Lena Rogström

Institutionen för svenska, flerspråkighet och språkteknologi, Göteborgs universitet

Modern lexikografi baseras idag på stora elektroniskt tillgängliga korpusar som kan analyseras med en mängd digitala verktyg. Ett sådant är SPARV (Borin et al. 2016) som möjliggör morfosyntaktisk analys av texter, vilket t.ex. är användbart vid kartläggning och studier av språkbruk på ord-, fras- och meningsnivå.

SPARV kan användas vid analys av äldre texter, men för text tillkommen före 1800-talet är den större variationen i stavning fortfarande ett problem, vilket begränsar analyser för t.ex. språkhistoriska ändamål.

I vår posterpresentation visar vi hur vi inom projektet *Digitalisering av Kungliga Vetenskapsakademiens Handlingar* (finansierat av Centrum för Digital Humaniora vid Hum.Fak., Göteborgs universitet) utnyttjar AI och teckenigenkänningsverktyg samt moderna lexikologiska metoder på vetenskaplig text från 1700-talet för att kartlägga framväxten av ett vetenskapligt ordförråd. Vi har t.ex. upprättat ordlistor i syfte att skapa en bild av det ordförråd som användes då och som skiljer sig ifrån dagens språk, antingen ortografiskt eller pragmatiskt. Genom att använda lexikala profiler kan graforden studeras ur ett frekvensbaserat perspektiv samt jämföras med ordlistor med modern svenska.

Vi analyserar också materialet i Språkbankens analysverktyg SPARV i syfte att kunna upprätta lexikaliska profiler över såväl specifika subgenrer som författare. I detta arbete är vårt mål att kunna visa hur delar av den ortografiska variationen kan kopplas till KARPs lexikon (Borin et al. 2012) och därigenom förbättra analysmöjligheterna också för texter från 1700-talet.

Dessa resultat kan utnyttjas för såväl historisk lexikografi (SAOB.2) som för vår kunskap om fackspråksetablering, vetenskapligt språkbruk och lexikalisk utveckling generellt sett.

Referenser

Lars Borin, Markus Forsberg, Martin Hammarstedt, Dan Rosén, Roland Schäfer, Anne Schumacher. 2016. Sparv: Språkbanken's corpus annotation pipeline infrastructure. *Proceedings of SLTC 2016*. The Sixth Swedish Language Technology Conference.

Lars Borin, Markus Forsberg, Leif-Jöran Olsson, and Jonatan Uppström. 2012. The open lexical infrastructure of Språkbanken. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul: ELRA. 3598-3602

Nyckelord: Fackordförråd; 1700-talstext; morfosyntaktisk analys; lexikal profil