

## SVALex: en andraspråksordlista graderad enligt CEFR nivåer

*Elena Volodina*

Institution för svenska, Göteborgs universitet  
elena.volodina@svenska.gu.se

*Ildikó Pilán*

Institution för svenska, Göteborgs universitet  
ildiko.pilan@svenska.gu.se

Ordlistor anpassade efter CEFR (Common European Framework of Reference, Ramverket) kan vara till mycket hjälp när man producerar andraspråksläromaterial och tester, men sådana ordlistor är svåra att producera. Även om Ramverket (COE 2001) har stimulerat initiativ med Reference Level Descriptions ([http://www.coe.int/t/dg4/linguistic/dnr\\_EN.asp?](http://www.coe.int/t/dg4/linguistic/dnr_EN.asp?)), har ännu inte det svenska initiativet tagit fart. För svenska har hittills funnits en Kelly-lista som har anmärkningar för användbarhet på nivåer enligt CEFR (GERS på svenska) (Volodina & Johansson Kokkinakis, 2012), men listan har inte validerats mot receptiva eller produktiva andraspråkskällor. Till skillnad från Kelly-listan som producerats baserat på webb-korpusar, har SVALex utgått ifrån läroböcker i svenska som andraspråk som används på CEFR-baserade kurser, och återspeglar därför den receptiva vokabulären som L2 elever möter under kurser på olika nivåer.

För att producera SVALex har vi använt COCTAILL, en korpus med läroböcker för svenska som andraspråk (Volodina et al., 2014). Annoteringen av COCTAILL inkluderar indelningen av material i texter avsedda för läsförståelse och allt annat material, bl.a. övningar, listor, exempel per nivå och kursbok. Dessutom, har lingvistisk annotation i form av ordklasser, grundformer, morfosyntaktiska och syntaktiska relationer automatiskt lagts till. Annotationen har möjliggjort ett urval av texter (läsförståelsematerial), där vårt fokus var på att samla frekvensinformationen på lemgramnivå, dvs. unika kombinationer av grundformer och ordklasser (t.ex. vara, substantiv).

SVALex innehåller 15,861 unika ingångar, 10% av dessa är flerordsenheter. På varje nivå används ett antal lemgram som elever stöter på för första gången, medan en hel del upprepas från nivå till nivå. Under konferensen kommer vi att fokusera på det ordförråd som finns i SVALex listan. Vi ska visa interaktiva sökningar, demonstrera några intressanta fall, samt koncentrera oss på ordförrådet i allmänna drag: Är polysema ord mest frekventa på nybörjarnivåer eller avancerade nivåer? När börjar man lära sig flerordsuttryck och partikelverb? Hur distribuerar sig ordklasserna mellan nivåerna? Blir orden längre ju mer avancerad språkinläraren blir? Vilket vokabulär hör till centralt versus periferärt på varje nivå och hur kan man bestämma detta? Vi kommer också att diskutera olika sätt att få fram en sådan lista med tanke på dess pålitlighet och omfång, och beskriva vår metod i korta drag. Man kan söka i och ladda ner listan från <http://cental.uclouvain.be/svallex/>

*Nyckelord:* svenska som andraspråk; frekvensbaserad ordlista; receptivt vokabulär.