**16:e Konference om Leksikografi i Norden (Lund, 27-29 april 2022)**

Peter Juel Henrichsen
Dansk Sprognævn
pjh@dsn.dk

## *Abstract*

# Det Centrale Ordregister

### Et indeks for det danske ordforråd - en gave til dansk sprogteknologi

Vi præsenterer det igangværende projekt *Det Centrale Ordregister*, samt et af projektets afleveringer, *COR-linkeren*. Projektets mål er at fremstille og distribuere et register kaldet COR (Det Centrale Ordregister) åbent for alle danske lemmaer og ordformer. COR's formelle design sikrer at (*i*) ethvert registreret lemma har et unikt COR-indeks (globally unique identifier); (*ii*) enhver bøjningsform afledt af et lemma *L* har et COR-indeks afledt af *L*'s indeks*;* (*iii*) det centrale danske ordforråd (repræsenteret af den danske retskrivningsnorm fastlagt i Lov Om Dansk Retskrivning) udgør COR's niveau 1, og i denne del er indekseringen komplet; (*iv*) alle øvrige danske lemmaer og ordformer kan indekseres i COR's niv. 2 og niv. 3 efter formelle kriterier. Et af de vigtigste formål med projektet - bortset fra at designe, konstruere og distribuere COR-indekset og dets relaterede ordressourcer - er at udvikle en COR-linker af høj lingvistisk kvalitet. Linkeren kan tage et tekstkorpus som input og afleverer hvert token annoteret med COR-indeks. Det COR-opmærkede output er, i sagens natur, disambigueret for homografi. Med linkerens hjælp bliver det derfor lettere for sprogteknologen at sammenlægge uafhængige korpora samt at søge information om de enkelte ordforekomster (PoS, udtale, betydning, emneklassifikation, sentiment m.v.) i danske ordbøger og databaser der er gjort COR-kompatible. Ved NFL16 præsenteres såvel COR-linkerens funktion som COR-indeksets overordnede struktur. Der gives også information om COR-projektets samlede afleveringer, der alle bliver publiceret i 2022-23. Naturligvis som open-source.