

Det Centrale Ordregister og ROhist

Thomas Widmann

Dansk Sprognævn

Det Centrale Ordregister (COR) tildeler unikke id-numre til alle lemmaer og ordformer på dansk. Grundordforrådets opdeling i lemmaer er baseret på *Retskrivningsordbogen* fra Dansk Sprognævn.

Dansk Sprognævn ROhist-projekt er en søgemaskine hvormed man kan sammenligne retskrivningerne fra 1872 til 2012. Det er en videreudvikling af det svenske SAOLhist-system.

I første omgang vil id-numrene i COR svare til den nyeste udgave af Retskrivningsordbogen (4. udgave fra 2012), men vi håber på at vi på sigt også kan tildele COR-numre til de historiske ordbøger i ROhist. Det vil foregå således at det samme nummer bruges hvis lemmaet er det samme – også hvis stavemåden har ændret sig. Et ord som *fråse* (RO2012) vil altså få samme COR-nummer som *frådse* (RO1996) og som *fraadse* i Dansk Haandordbog (1872). Dette vil simplificere implementeringen af ROhist betydeligt. COR indeholder også mekanismer som gør det muligt at vise diakrone ændringer – især sammenfald og opsplittings af lemmaer.

COR-opmærkninger af historiske ordbøger har også mange andre anvendelser. Man vil fx kunne (1) ændre stavemåden af en COR-opmærket tekst fra en ortografi til en tidligere eller senere, (2) opdatere tosprogede ordbøger til nyeste udgave af retskrivningen, og (3) opdatere datalingvistiske værktøjer som fx taggere (hvilket vil understøtte opmærkningen af historiske tekstkorpusser).

Til sidst vil jeg se på muligheden for engang at forbinde dansk og norsk bokmål i værktøjer som ROhist. De to deler som bekendt et ortografisk udgangspunkt, så når vi opmærker de historiske ordbøger med COR-id-numre, vil det blot kræve et norsk parallelprojekt for kunne forbinde de to sprogs retskrivninger. Hvis fx det moderne *fråse* har samme COR-nummer som *fraadse* i sidste fælles ordbog, vil man kunne tage det tilsvarende norske id-nummer og se at de moderne former dér er *fråtse* og *fråsse*.

Nøgleord: ordbøger; sprogteknologi; sproghistorie