

## ***DK-CLARIN Statusrapport***

### ***Aflevering nr. D 4.2.2.1b***

Arbejdspakke, nummer og navn	4.2.2
Aflevering, navn	Frekvensbaseret selektion af ordforrådet
Afleveringsfrist	T18
Afleveringsdato	30. juni 2009
Projektleder for arbejdspakken	Anna Braasch
Deltagere	Lene Offersgaard og Sussi Olsen
Evt. tilhørende dokumenter på hjemmesiden	

Selektion af lemmakandidater var den første arbejdsopgave i WP4.2.2 med henblik på at sammenføre den SprogTeknologiske Ordbases (STOs) morfologiske og syntaktiske beskrivelser af ord med de samme ords semantiske beskrivelser i DanNet.

Vi har foretaget den planlagte frekvensbaserede selektion af fællesmængden af ordforrådet i STO og DanNet med henblik på udvælgelse af materiale til sammenkædningens første runde (til aflevering: T28).

Der er selekteret 8.500 zerovalente substantiver som kandidater for automatisk sammenkædning i overensstemmelse med beskrivelsen i rapporten (aflevering 4.2.2.1a). Lemmalisten ligger på CST's interne server.