

## Rapport om projekt

# OrdRet

- en it-baseret stavekontrol

Udviklet af

Mikro Værkstedet A/S

GrammarSoft Aps

Dansk Videnscenter for Ordblindhed

Rapport om OrdRet - en it-baseret stavekontrol  
© Konsortiet OrdRet, 2005

Redaktion: Birgit Dilling Jandorf

ISBN: 87-990274-1-0

OrdRets logo er udviklet af grafikeren Soffi Beier.

Projektet er støttet økonomisk med midler fra:



It- og Telestyrelsen

Egmontfonden



Sygekassernes Helsefond



Undervisningsministeriets Tips - og Lottomidler

## Forord

Denne rapport redegør for udviklingen af den it-baserede stavekontrol OrdRet. Rapporten beskriver projektets livsforløb, udviklingen af selve programmet og de overvejelser, der ligger bag de sproglige resurser i stavekontrollen. Her er der således mulighed for at få at vide, hvordan stavekontrollen søger efter fejlord, og hvilke forslag den kommer med. Der er oplysninger om, hvordan programmet fungerer, og man kan læse om resultaterne af de forskellige afprøvninger. Endelig kan man få et indblik i, hvad vi har af ideer til en videreudvikling af programmet.

Mange mennesker har været involveret i arbejdet bag OrdRet. Professor Carsten Elbro og lektor, ph.d. Holger Juul fra Københavns Universitet og lektor, ph.d. Elisabeth Arnbak, Danmarks Pædagogiske Universitet har bistået med værdifuld indsigt om stavning. Lektor, ph.d. Peter Juel Henriksen fra Institut for Datalingvistik Handelshøjskolen i København har udtrukket en liste af forvekslelige og næsten forvekslelige ordpar, der har dannet grundlag for udviklingen af eksempelsætninger. En stribe af studentermedarbejdere har medvirket til, at eksempelsætningerne i programmet er udvalgt med omhu og formuleret med sproglig akkuratess. Tak til Sille Villain Meulengracht, Trine Thyrsted, Marie Poulsen Larsen og Heidi Ahlbom. I afprøvningen af programmet har vi allieret os med kompetente konsulenter og speciallærere Margit Gade, Heidi Thomsen og Leif Petersen. Tak for jeres gode råd.

På projektets vegne vil jeg gerne takke It- og Telestyrelsen og styregruppen for Kompetencecenter It for alle, Undervisningsministeriet (Tips- og Lotto), Egmontfonden og Sygekassernes Helsefond, der troede på vores projekt og gav os midler.

Endelig vil jeg gerne takke de mange ordblinde børn, unge og voksne og deres lærere, der har medvirket med deres tekster og har afprøvet denne stavekontrol.

Denne rapport kan downloades fra hjemmesiden: [www.ordret.com](http://www.ordret.com)

Birgit Dilling Jandorf  
Projektleder

Dansk Videnscenter for Ordblindhed  
Juni 2005

## Indhold

Forord.....	3
Baggrund.....	5
Partnere og deres ansvar i OrdRet .....	5
Formål.....	6
Projektets livsforløb.....	6
Udviklingen af programmet OrdRet.....	7
Planlægning og konceptudvikling.....	7
Programudvikling og sprogforskning .....	8
Moduler .....	8
Afprøvning og test .....	9
Kommunikation og samarbejde .....	9
Overblik over eksisterende stavekontroller.....	11
Dansk stavekontrol.....	11
Nordiske stavekontroller .....	12
Indsamling af ordblindes stavefejl .....	12
Teksttyper .....	13
Resultater.....	14
Udvikling af fejltypologi .....	15
De sproglige resurser i OrdRet .....	17
Kontekstfrie resurser og leksikografi .....	17
Listekontrol og komposita.....	17
Fejldatabaser .....	17
Ordlighed.....	18
Forslagsprioritering.....	18
Fejl i korrektstavede ord.....	18
Grammatiske, kontekstbaserede resurser .....	19
Evaluerings af OrdRets sproglige resurser .....	21
Ekstern afprøvning af OrdRet .....	23
Præsentation af programmet.....	24
Opsamling og evaluering af projektet.....	27
Forventninger til kommercielle resultater af projektet .....	27
Perspektiver .....	28
Referencer .....	29




## Baggrund

En relativ stor del af den danske befolkning har svært ved at læse, skrive og stave. Mange er så hårdt ramt, at det kniber med at mestre ny teknologi og leve op til de krav om læse- og stavefærdigheder, som møder dem i hverdagen, på arbejdet og i uddannelsen. Ordblindes stavevanskeligheder er i særdeleshed sejlivede og langvarige at få bugt med. Ordblindes centrale problem er at mestre koblingen mellem bogstaver og sproglyde. Mange ordblinde betragter stavning og skrivning som en meget arbejdskrævende og uoverskuelig proces, som de meget sjældent har succes med. På grund af stavefejlene kan de ofte slet ikke læse, hvad de har skrevet, og når de skriver, undgår de ord, som de ikke kan stave.

I nærværende projekt er målgruppen primært ordblinde, men der er betydeligt flere i Danmark, der har problemer med retskrivning. Et realistisk anslag er, at ca. 12 % af befolkningen har så store problemer med retskrivning, at de har behov for mere end, hvad almindelige stavekontroller kan tilbyde. Hertil kommer tosprogede, hvis behov ligeledes rækker ud over almindelige stavekontroller. Vi tror derfor, at vi kan udvikle en stave- og skrivehjælp, som rigtig mange vil få glæde af.

## Partnere og deres ansvar i OrdRet

Projektet består af tre partnere, der udgør OrdRet-konsortiet.

-  **Mikro Værkstedet A/S** har med direktør Niels Askær, udviklingschef Torben Worm og programudviklerne Andreas Bang Petersen og Casper Wittorff været ansvarlige for programmets kravspecifikation, brugergrænseflade, funktionalitet og integration med Windows og VISL. De har ligeledes været ansvarlige for integration af syntetisk tale, installation og afprøvning af programmet.
-  **GrammarSoft Aps** har med sprogforskerne John Dienhart, Eckhard Bick, samt programmør Tino Didriksen været ansvarlig for definitionsafgrænsning og leksikografi, integration med DanGram, produktion af Constraint Grammar-moduler, udvikling af algoritmer og kvalitetssikring.
-  **Dansk Videnscenter for Ordblindhed** har været ansvarlig for projektledelse, dataindsamling, udvikling af typologi, kategorisering, udvikling af sproglige tekster til stavehjælpen, offentlig lancering og rapportering af projektet.

## Formål

Projektets formål er at udvikle en it-baseret stave- og skrivehjælp og teste dens afgrænsning, sproglige støtte og funktionalitet i forhold til målgruppen af ordblinde børn, unge og voksne. Vores ambition er, at den danske stavekontrol kan udstyre brugeren med et pædagogisk værktøj, der udover at have en støttende funktion også skal have en lærende funktion.

Den støttende funktion skal bestå i, at programmet kan identificere det fejlstavede ord og læse det op med syntetisk tale og dernæst give et antal valgmuligheder for korrekt brug af ordet i forskellige kontekster.

Den lærende funktion skal bestå i, at programmet indeholder en mulighed for, at brugeren selv kan bearbejde og rette det fejlstavede ord i samspil med en computergenereret liste af forslag. Der skal ligeledes være mulighed for at lægge nye ord ind i programmet. Dette vil give mulighed for at inddrage ord, som brugeren har behov for - samt anvende disse ord i en pædagogisk proces.

## Projektets livsforløb

Allerede i 2000 blev vi inspireret til at få udviklet en ny dansk stavekontrol - primært designet til ordblinde. Det var bl.a. ved en præsentation af den svenske stavekontrol "Stava Rätt", at videnscenteret tog kontakt til Mikro Værkstedet og fortalte om "Stava Rätt". Der blev arrangeret et møde på Syddansk Universitet med deltagelse af direktør Niels Askær og Jens Erik Rasmussen fra Mikro Værkstedet og sprogforskerne John Dienhart og Eckhard Bick fra GrammarSoft. På mødet stiftede vi bl.a. bekendtskab med forskningen bag constraint grammar. Mødet blev grundsten for et projektsamarbejde, der skulle strække sig over de næste mange år.

Desværre måtte vi undervejs i projektet tage afsked med John Dienhart. Han døde efter kort tids sygdom. Projektets videre skæbne blev diskuteret meget nøje, og alle partnere blev enige om, at projektet skulle fortsætte – i John Dienharts ånd.

Som udgangspunktet for samarbejdet udarbejdede Dansk Videnscenter for Ordblindhed en kort projektbeskrivelse. Dernæst fulgte en række møder om forudsætningerne for projektet, afklaring af de gensidige forventninger og forpligtelser.

Videnscentret udarbejdede derefter en egentlig projektbeskrivelse med et forslag til ansvarsfordeling, en tids- og aktivitetsoversigt og et budget. Så begyndte en intensiv ansøgningsperiode med udsendelse af i alt 25 ansøgninger.

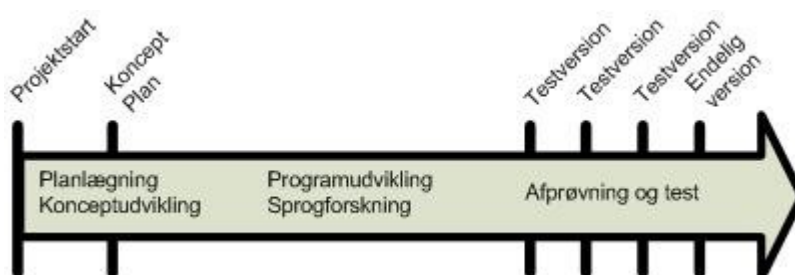
Heldigvis var der fire af dem, der troede på os og vores projekt, og i slutningen af december 2003 kom de sidste penge i hus.

Undervejs fik vi beskrevet en samarbejdsaftale, der definerede parternes pligter og rettigheder under udviklingsarbejdet og i forhold til en eventuel efterfølgende kommerciel udnyttelse. Samarbejdsaftalen blev underskrevet den 11. juni 2004.

Den overordnede koordinerende aktivitet for projektet har været de relativt ofte afholdte styregruppemøder, hvor beslutningsdygtige repræsentanter for projektets parter har været til stede sammen med dele af den operationelle del af projektet. Dette har betydet en meget kort "kommandovej" og effektive beslutninger i projektet, samt en god kommunikation mellem parterne i hele projektet. Der har været afholdt 15 effektive styregruppemøder.

## Udviklingen af programmet OrdRet

Dette afsnit beskriver overordnet udviklingen af programmet OrdRet og tager udgangspunkt i processen omkring udviklingen samt de enkelte produkter (dokumenter og programdele), der indgår i OrdRet.



Figur 1: Projektets faser

Udviklingen har fulgt en proces, der lineært set indeholder tre store faser, planlægning og konceptudvikling, programudvikling og sprogforskning og afprøvning og test som vist på figur 1.

### Planlægning og konceptudvikling

Den første fase var en planlægnings- og konceptudviklingsfase, hvor der blev arbejdet intensivt med at lave det første design af brugergrænsefladen samt at fastlægge de væsentligste funktionaliteter i OrdRet.

Brugergrænsefladen blev udarbejdet på et antal diskussionsmøder mellem projektets parter og resultaterne af disse diskussioner blev skrevet ind i konceptdokumentet. Konceptdokumentet er det overordnede dokument, der

beskriver produktet. Det danner udgangspunkt for alle senere dokumenter og beslutninger i projektet.

På baggrund af konceptdokumentet blev der skrevet en egentlig kravspecifikation, der i detaljer beskriver hele produktet inklusiv alle funktioner.

Formålet med kravspecifikationen er at beskrive produktet så detaljeret, at der kan programmeres ud fra det. Der må ikke være uafklarede punkter i den. Kravspecifikationen blev primært skrevet af Mikro Værkstedet i en meget tæt dialog med projektets to andre parter, så alle uklare detaljer blev diskuteret på så tidligt et tidspunkt som muligt.

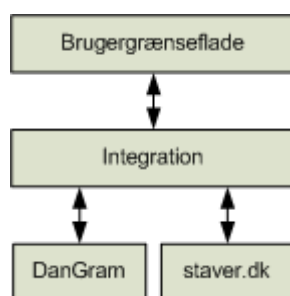
Konceptdokumentet og kravspecifikationen er de primære input til planlægningsprocessen og opdelingen af arbejdsopgaver mellem de tre projektparter. På baggrund af disse dokumenter blev der udarbejdet en projektplan, der blev anvendt gennem hele projektet.

## Programudvikling og sprogforskning

Den næste fase bestod af den lingvistiske udvikling og udviklingen af selve programmet OrdRet. Den lingvistiske udvikling omfattede både indsamling og analyse af ordblindes stavfejl, leksikografi og frekvensanalyser, samt udvikling af lingvistiske analyse- og genereringsredskaber, lydskrivnings- og komparationsalgoritmer.

### Moduler

Teknisk set består OrdRet af tre moduler (se Figur 2), der er forbundet gennem et integrationsmodul. Denne struktur betyder, at der er gode muligheder for, at der kan gå noget galt i denne integration, da der er forskellige personer involveret i de forskellige dele. Det blev derfor besluttet at udvikle OrdRet ved hjælp af en inkrementel proces, hvor små dele af programmet udvikles og afprøves i et samlet hele frem for at lave delene først og samle dem bagefter.



Figur 2: OrdRets moduler



De tre dele, som OrdRet består af er en *brugergrænseflade*, *DanGram* (grammatikmodul, ordkomparation, kontekstuel fejltypebestemmelse og forslagsvægtning) og dele af *staver.dk* (ordinformation).

Brugergrænsefladen var den første del, der blev udviklet, idet vi med en fungerende brugergrænseflade hurtigt kunne vise programmet frem for de eksterne afprøvere (se side 23) og dermed kunne projektet hurtigt få feedback på, om det valgte design kunne anvendes til målgruppen. Brugergrænsefladen blev udviklet i et tæt samarbejde mellem videnscenteret (som i denne sammenhæng repræsenterer slutbrugerne) og Mikro Værkstedet som tekniske eksperter.

DanGram er selve det grammatiske analysemodul, hvor den grammatiske analyse af brugerens tekst foregår. Udviklingen af dette modul har været en stor udfordring, idet det ikke tidligere har været prøvet at lave sætningsanalyser på ufuldstændige (ikke korrektstavede) sætninger. Det har derfor involveret en betydelig forskningsindsats at udvikle dette modul, og det er netop i udviklingen af dette modul, at det indsamlede korpusmateriale har været anvendt intensivt.

Staver.dk er Mikro Værkstedets ikke-kontekstbaserede stavekontrol. En delmængde af denne stavekontrol indgår som en del af kvalificering af det input, som DanGram arbejder på. Staver.dk leverer også bøjningslister og anden information på ordniveau til slutbrugeren. Den sidste del af OrdRet er den del, der binder de tre hoveddele sammen. Denne del er primært udarbejdet af GrammarSoft i tæt sparring med Mikro Værkstedet.

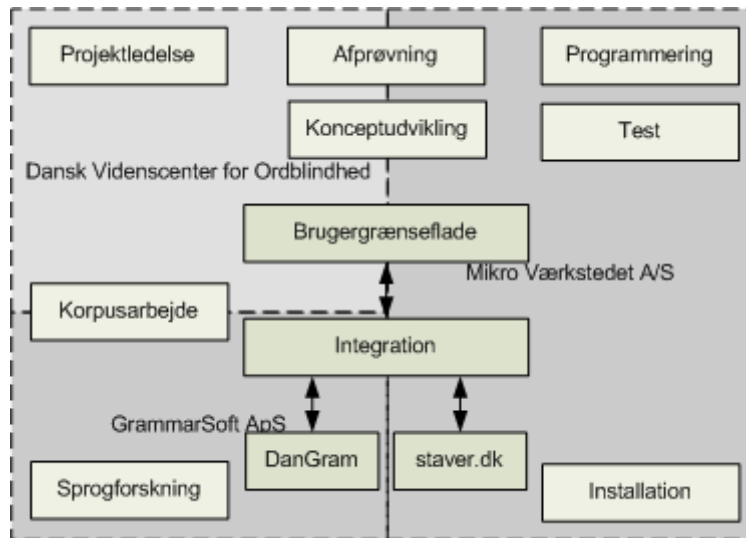
## Afprøvning og test

Den sidste fase var test- og afprøvningsfasen, der involverer såvel eksterne domæneeksperter samt interne tekniske testere og en evaluering mod en håndopmærket "guld-standard"-tekst. De eksterne testes involvering er beskrevet side 23.

Den interne test af OrdRet består i at teste, at produktet teknisk set er i orden, dvs. at produktet kan installeres på de ønskede platforme, og at alle de i kravspecifikationen beskrevne funktioner er implementeret og fungerer efter hensigten. Performanseevalueringen skulle slå fast, hvor god OrdRet er til at fange ordblindefejl og til at vægte rettelsesforslag.

## Kommunikation og samarbejde

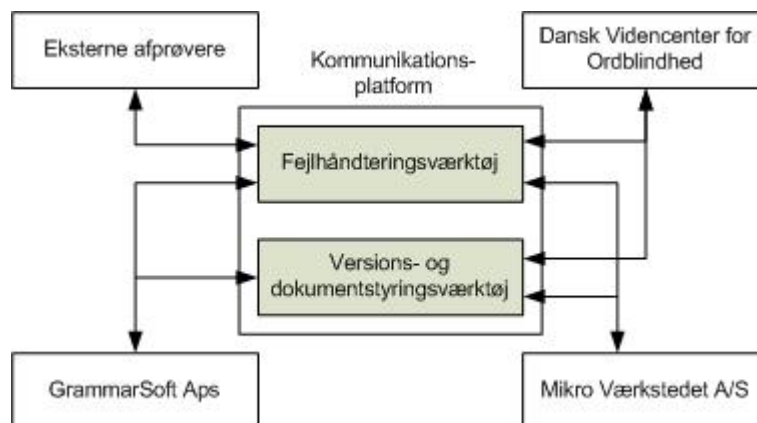
Ud over den rent tekniske/logiske opbygning af selve programmet og det deraf følgende arbejde er der en del andre aktiviteter og samarbejde i projektet. Disse aktiviteter er vist på Figur 3: Aktiviteter



Figur 3: Aktiviteter

Figuren viser både de opgaver, der er udført i samarbejde (vist på kanterne af kasserne), f.eks. korpusarbejde og konceptudvikling, samt de opgaver, der har været udført af de enkelte parter.

For at gennemføre samarbejdet på en effektiv måde på trods af den geografiske spredning mellem projektets parter, valgte vi at indføre en kommunikationsplatform bestående af to værktøjer – foruden e-mail, telefon, møder og anden kommunikation. De to værktøjer var et *Fejlhåndteringsværktøj* (*Bugzilla*) og et *versions- og dokumentstyringsværktøj* (*CVS*) som vist i Figur 4.



Figur 4: Bugzilla og CVS

Fejlhåndteringsværktøjets funktion var at skabe en struktureret opsamling af fejl og forbedringsforslag på tværs af såvel projektets parter som eksterne afprøvere. Det skal dog her nævnes, at fejlhåndteringsværktøjet viste sig at være for

komplekst for de eksterne testere. Der blev senere udviklet en forenklet adgang til dette værktøj til fremtidig ekstern test. De fejl og forbedringsforslag, der er indrapporteret, har gennem hele projektet været anvendt til at planlægge og prioritere ud fra (sammen med den øvrige projektplan).

Formålet med versions- og dokumentstyringsværktøjet var at skabe en fælles platform for at dele dokumenter, grafik og kildetekst i projektet på en sådan måde, at alle havde adgang til så mange informationer som muligt. Desuden var det væsentligt, at det var muligt for alle at rette i de forskellige dokumenter og stadig være sikre på, at de versioner, der blev anvendt i projektet var de korrekte.

## Overblik over eksisterende stavekontroller

### Dansk stavekontrol

I Danmark findes ikke en stavekontrol, som er særligt udviklet til at imødekomme de stavfejl, som ordblinde laver. Langt de fleste mennesker – med og uden specifikke retskrivningsvanskeligheder – anvender stavekontrollen i tekstbehandlings-programmet MS Word. Gennem projektets forløb er der sket væsentlige forbedringer i stavekontrollen i Word.

En begrænsning ved stavekontrollen i Word er, at den ofte ikke har et kvalificeret bud på et ordforslag til et givet fejlstavet ord. Stavekontrollen er således ikke er god nok til at "gætte" brugerens intenderede ord. De ordforslag, stavekontrollen tilbyder, er ikke hensigtsmæssigt prioriteret, hvilket betyder, at der ikke er sikkerhed for, at det intenderede ord fx ikke først er forslag nummer otte i rækken af mulige ord.

Konsekvensen er, at brugeren stilles over for en lang liste af valgmuligheder i identifikationen af det intenderede ord. Det er en svær opgave for mange ordblinde, fordi ordene i listen ofte lydligt, men måske også grafisk ligner hinanden. Årsagen til dette er bl.a., at Word ikke prioriterer sine ordforslag i forhold til konteksten, hvilket medfører, at der ofte gives irrelevante alternativer, det vil sige ord, der ikke passer ind i den syntaktiske kontekst.

En anden begrænsning ved stavekontrollen i Word er, at rigtige navneord og specielle fagudtryk markeres som fejl. Word kender fx ikke ord som *videnscenter* og *fejlord*. Hvis man som bruger af stavekontrollen til trods for den røde fejlmarkering har en formodning om, at ordet alligevel er korrekt stavet, så har man ingen mulighed for at få tjekket, om formodningen er korrekt.

Word kan ikke identificere og skelne mellem fejlstavede ord, som er homofone, det vil sige enslydende, men også næsten enslydende ord skelnes ikke. Både enslydende og næsten enslydende ord forveksles ofte, men fejlmarkeres

ikke i Word, fordi der er tale om korrekte stavede ord. Et eksempel på let forvekslelige ord er *bager - bæger - bærer - bære*. Konsekvensen af, at Word ikke fejlmærker ord af denne type, er, at en sætning som *Du kan købe kager og brød hos en bære* vil gå glat igennem uden fejlmærkning. Årsagen til denne fejl er stadig, at Word ikke tager konteksten i betragtning, det vil sige, at Word ikke tager hensyn til den syntaktiske sammenhæng, som det aktuelle ord indgår i.

I Word kan man ikke foreslå korrekt stavning ved stavefejl, der ikke ligger lige op ad den intenderede stavemåde. Denne begrænsning medvirker til at mindske anvendeligheden af stavekontrollen for ordblinde, fordi de ofte begår stavefejl, som både grafisk og fonetisk ligger langt fra det intenderede ord. Et eksempel er fejlordet *blitter* for ordet *billetter* i sætningen *De gik ud og købte blitter*.

Endelig er der ikke mulighed for oplæsning af ordforslag og hjælpetekster i Word. Dette er ganske problematisk i forhold til gruppen af ordblinde, der ofte, udover stavevanskeligheder, også har læsevanskeligheder.

## Nordiske stavekontroller

I år 2000 stiftede Dansk Videnscenter for Ordblindhed bekendtskab med den svenske stavekontrol *Stava Rätt*. Denne stavekontrol er særligt udviklet til ordblinde. Den er baseret på typiske ordblindedefejl og kan med en meget høj grad af præcision finde og korrigere de stavefejl, ordblinde laver. *Stava Rätt* er i en svensk undersøgelse blevet sammenlignet med stavekontrollen i Word, og denne sammenligning viser, at hvor *Stava Rätt* finder og retter knapt 80 % af de stavefejl, som ordblinde laver, så klarer Word at finde og rette 30 – 35 % af de stavefejl, som ordblinde laver.

Tilsvarende har det norske projekt *Norsk tekstbehandling for dyslektikere – Lingvistisk basert staveverktøy*, forkortet *LingDys*, dokumenteret samme resultater både med hensyn til genfindning og præcision.

I vore nordiske nabolande findes altså allerede stavekontroller, der imødekommer de begrænsninger, som ordblinde og andre med retskrivningsvanskeligheder møder. Ambitionen for OrdRet har derfor været, at udvikle en dansk stavekontrol til ordblinde, der er mindst lige så god – og måske endnu bedre end de nordiske.

## Indsamling af ordblindes stavefejl

I Danmark er der i nyere tid ikke gennemført en større systematisk indsamling og analyse af ordblindes stavefejl. Nutidig viden om danske ordblindes stavning er derfor ganske begrænset og præget af mere eller mindre usystematiske læsepædagogiske erfaringer.

Videnscenteret gik derfor i gang med en omfattende indsamling af tekster skrevet af ordblinde. Formålet med indsamlingen var at få en bred viden om ordblindes stavning; derfor blev der indsamlet stavfejl fra ordblinde i alle aldre, dvs. børn, unge og voksne.

I alt blev der taget kontakt til 36 amtslige og kommunale institutioner for ordblinde i Danmark. Alle institutioner var registreret som undervisningsinstitutioner for ordblinde i videnscentrets informations- og adressedatabase.

Første kontakt bestod i et brev til institutionernes forstandere, hvor projektet blev beskrevet, og vi anmodede forstanderne om at finde lærere, der ville indsamle skriftligt materiale fra deres elever. Når vi fra hver institution fik tilsagn om, at et antal lærere og deres elever ville deltage i projektet, sendte vi de enkelte lærere et udførligt brev med nøje instruktion i, hvordan teksterne skulle indsamles.

## Teksttyper

Der blev indsamlet to typer af tekster; frie skriftlige formuleringer og diktater. Et krav til både de frie skriftlige formuleringer og diktaterne var, at de skulle være produceret uden nogen form for hjælpemidler. På alle diktater og frie skriftlige formuleringer blev deltagerens køn og fødselsår påført.

Fri skriftlig formulering er i denne sammenhæng tekst, der er produceret uden nogen form for bundet oplæg. Emnet for den skriftlige formulering var således frit, men lærerne blev dog opfordret til at foreslå et eller flere emner, så deltagerne, der ikke selv havde noget på hjerte, kunne få inspiration til at gå i gang med den skriftlige produktion. Ligeledes blev deltagerne opfordret til blot at skrive løs uden hensyntagen til eventuelle fejl. Frie skriftlige formuleringer blev inddraget, fordi det er i forbindelse med produktion af denne form for skriftligt materiale, at ordblinde vil komme til at anvende stavkontrol. Ligeledes er fri skriftlig formulering en naturlig måde at producere noget skriftligt på.

At benytte frie skriftlige formuleringer giver i modsætning til diktater en god mulighed for at undersøge, hvordan deltagerne staver i en naturlig skrivesituation. Deltagerne tilskyndes således til at trække på deres eget lingvistiske repertoire. Denne metodiske tilgang tillader undersøgelse af en bred vifte af forhold vedrørende stavning. Deltagerne skriver deres egne ord, ikke ord, som vi gerne ville undersøge. Ved at anvende frie skriftlige formuleringer sikres, at resultaterne ikke skævvrides mod eller væk fra en bestemt stadiebeskrivelse af stoveudviklingen enten ved at inkludere eller ekskludere ord baseret på en bestemt stadiebeskrivelse af stoveudviklingen. Datamaterialet tillader os at undersøge, hvordan deltagerne naturligt skriver ord og laver stavfejl.

Når man indsamler data i form af frie skriftlige formuleringer skal man være opmærksom på, at deltagerne måske undlader at benytte ord, som de har svært ved at stave. En anden begrænsning ved denne form for data er, at der ingen kontrol er over variable så som ordenes frekvens og længde. Dette er faktorer, som har velkendt indflydelse på stavning.

Et krav til de frie skriftlige formuleringer var, at de skulle være skrevet på computer. Dette krav levede alle tekstbidrag dog ikke op til, hvilket betød at vi efterfølgende selv måtte indskrive bidragene. Denne procedure medfører dels fare for at lave tastefejl og dels fare for subjektiv vurdering ved tydning af skriverens håndskrift.

Diktater blev inddraget af flere grunde. Diktater indeholder ofte ord, som er vanskelige at stave, og som derfor kan fremprovokere diverse stavefejl. Endvidere var det forventeligt, at deltagerne i de frie skriftlige formuleringer udelukkende ville benytte sig af ord, som de ikke havde vanskeligt ved at stave. Ved at inddrage diktater, forventede vi, at det samlede ordmateriale ville rumme ord, som deltagerne ikke nødvendigvis selv ville have produceret på deres aktuelle stadie af staveudviklingen. I inddragelsen af diktater kunne vi således udvide mængden af forskellige ord og også få flere fejlstavede ord i det samlede materiale. Der var ikke tale om bestemte på forhånd udvalgte diktater. Grunden hertil var, at vi vurderede, at flere forskellige diktater ville føre til flere forskellige typer fejl. Og da ønsket var så mange fejl som muligt, var det hensigtsmæssigt at samle fejl fra så mange forskellige diktater som muligt. De diktater, vi har inddraget i undersøgelsen, er traditionelle orddiktater. Traditionelle orddiktater er diktater, hvor et diktatorord bliver præsenteret for deltageren i en sætning. Deltageren skal herefter nedskrive diktatorordet. Vi modtog næsten alle diktater i håndskreven form og måtte herefter selv indskrive dem. Igen skal det nævnes, at denne procedure medfører fare for at lave tastefejl samt fare for subjektiv vurdering ved tydning af skriverens håndskrift. Vi modtog endvidere diktatforlæg til alle diktater, således at vi kunne se, hvad et eventuelt fejlstavet ord skulle have været.

## Resultater

Vi modtog i alt 1244 bidrag, heraf 646 frie skriftlige formuleringer og 598 diktater. Vi valgte at inddele deltagerne i tre grupper. Børn er 7-12 år, unge er 13-17 år, og voksne er 18-62 år. Deltagerne fordeler sig aldersmæssigt således: 105 børn, 673 unge og 313 voksne.

Vi har ikke mange børn med i undersøgelsen. Årsagen til at vi ikke har yngre børn med er, at stave- og skriveintensiteten i førskolealderen er minimal, og vi har derfor ikke kunnet få deltagende førskolebørn. Årsagen til at vi

har så få børn med i det hele taget er, at mange børn først identificeres og diagnosticeres som ordblinde sent i skoleforløbet (Haven & Nielsen, 2004).

Samtlige tekster er lagt i en database, som vi har udviklet specielt til formålet. I databasen er alle tekster, dvs. hvert ord i hver tekst, registreret med alder og køn og er forsynet med et identifikationsnummer, så vi fra et givet ord altid kan finde tilbage til den tekstsammenhæng, det indgår i, og til oplysninger om alder og køn på den deltager, der har produceret teksten, hvori ordet forekommer. I alle frie skriftlige formuleringer og diktater er stavfejl identificeret og noteret i databasen.

Til at identificere stavfejl i frie skriftlige formuleringer har vi udviklet et sæt retningslinjer for rettelserarbejdet, så måden hvorpå fejl blev identificeret kunne gøres konsekvent for alle 1244 tekster. Retningslinjerne kan virke "firkantede", men i hvert enkelt tilfælde har vi været nødt til at trække en grænse, så identifikationen af en given fejl ikke skulle bero på et subjektivt skøn. I nogle tilfælde har vi ikke med tilstrækkelig stor sandsynlighed kunnet identificere det intenderede ord. I disse tilfælde har vi registeret ordet som uforståeligt. Retningslinjerne er udarbejdet i fællesskab, men for at sikre konsekvens i måden, hvorpå teksterne blev rettet, har kun én person været involveret i rettelserarbejdet. Vi har valgt at registrere alle ord, både fejl og rigtigt stavede ord, for at få mulighed for at se på, hvilke ord ordblinde har svært ved at stave og hvilke ord, ordblinde rent faktisk kan stave. Til at identificere fejl i diktaterne brugte vi de tilhørende diktatforlæg.

Vi har naturligvis også registreret alle rigtigt stavede ord således, at vi dels får viden om, hvilke ord deltageren ikke kan stave og dels, hvilke ord deltageren faktisk kan stave. Dette er en ganske værdifuld information, når vi ønsker at få indsigt i, hvilke ord der volder ordblinde særlige vanskeligheder.

I de frie skriftlige formuleringer er der skrevet i alt 113.616 ord, heraf er 21.539 fejlstavede ord. I diktaterne er der skrevet i alt 21.160 ord, heraf er 7.770 fejlstavede ord. I indsamlingen har vi altså i alt indsamlet 134.776 ord, heraf i alt 29.309 fejlord, hvoraf 6.090 er forskellige ord.

## Udvikling af fejltypologi

Til brug for analyse af de indsamlede stavfejl har vi udviklet en fejltypologi, og ud fra denne er alle stavfejlene analyseret. På baggrund af et (begrænset) antal undersøgelser af ordblindes stavning (Nelson 1980; Bruck 1988; Moats 1995; 1996; Lennox & Siegel, 1998; Sawyer et al.1999) og viden om det danske skriftsystem har vi opstillet følgende 14 fejltyper. Til selve fejlanalysen er der udviklet et sæt meget detaljerede retningslinjer, som ikke bringes her.

<b>Fejltype</b>	<b>Beskrivelse</b>	<b>Eksempel</b>
Lille > stort	Lille begyndelsesbogstav, som skulle have været stort.	<i>Adam &gt; adam</i>
Stort > lille	Stort begyndelsesbogstav, som skulle have været lille.	<i>januar &gt; Januar</i>
Særskrivning	Ét ord er skrevet i to eller flere dele.	<i>Billedanalyse &gt; Billede analyse</i>
Sammenskrivning	Ét ord, som skulle være skrevet i to ord.	<i>brug for &gt; brofor</i>
Lyd	Et ord som lydligt ligner det intenderede ord.	<i>adskillige &gt; adskelige</i>
Længde 1 > 2	Et ord med enkeltkonsonant, som skulle være skrevet med dobbelkonsonant.	<i>læsse &gt; læse</i>
Længde 2 > 1	Et ord med dobbelkonsonant, som skulle være skrevet med enkeltkonsonant.	<i>læse &gt; læsse</i>
Leksikalsk fejl	Et andet korrekt stavet ord end det intenderede.	<i>at &gt; og</i>
Rod	Et ord, hvori der er en fejl i roden.	<i>synge &gt; sønge</i>
Rod kompositum	Et sammensat ord, hvori der er en fejl i en eller flere rødder.	<i>atombombe &gt; andtom bombe</i>
Præfiks	Et ord, hvori der er en fejl i præfikset.	<i>forstå &gt; fårstå</i>
Suffiks	Et ord, hvori der er en fejl i suffikset.	<i>afhængigt &gt; afhængit</i>
Bøjning	Et ord, hvori der er en fejl i bøjningsendelsen, eller der er tilføjet en bøjningsendelse.	<i>aberne &gt; abener</i>
Andet	Et ord, som ikke kan placeres under de øvrige kategorier.	<i>gøde &gt; gøvede</i>

Tabel 1: Fejltypologi

Udover at alle stavfejl er inddelt efter fejltype, er alle stavfejl inddelt efter ordklasse. Vi har inddelt efter følgende ordklasser: adjektiv, adverbium, artikel, blandet (to ord af hver sin ordklasse), forkortelser, infinitivmærke, konjunktion, lydord, pronomen, proprium, præposition, substantiv, talord, udråbsord og verbum.

Alle stavfejl og analyser ligger som en sproglig resurse i OrdRet. Desværre har det indenfor projektets rammer ikke været muligt at udføre detaljerede analyser på den store base af ordblindes stavfejl. En detaljeret analyse vil dog blive gennemført i andet regi, og forhåbentligt vil resultaterne af denne analyse kunne medvirke til at kvalificere OrdRets anvendelse til ordblinde yderligere.



## De sproglige resurser i OrdRet

Udviklingen af de sproglige resurser i OrdRet er blevet varetaget af konsortiets lingvistiske arm, GrammarSoft. Med udgangspunkt i en Constraint Grammar-baseret tagger/parser for dansk (Bick, 2001) og tilhørende leksikografiske resurser er der skabt et modulært system, der både indeholder traditionelle stavekontrolkomponenter, lydskrivning, kontekst-drevne grammatikker til prioritering af rettelsesforslag, grammatisk fejlanalyse og generering af bøjningsformer. Arbejdet har været forskningsbaseret og involveret bl.a. leksikografi, udvikling af komparationsalgoritmer og anden programmering, korpuslingvistik, statistisk evaluering samt, ikke mindst, skrivning og testning af Constraint Grammar-regler<sup>1</sup>.

### Kontekstfrie resurser og leksikografi

#### Listekontrol og komposita

Rygraden i de fleste traditionelle stavekontrolprogrammer består af en liste af i normsproget accepterede ordformer. Et standardprogram vil ganske enkelt definere en fejl som et ord der ikke figurerer på listen. For dansk kan denne metode medføre mange falsk positive fejlmeldinger, fordi sproget forholdsvis frit sammensætter nye komposita, og OrdRet har derfor dels inddraget tekstkorpora til at opbygge et leksikon med sjældne komposita, dels indbygget en morfologisk analyse, der kan genkende ukendte ord som mulige komposita snarere end "røde" fejl<sup>2</sup>. Et andet problem med en ren listekontrol er, at listen vil acceptere de to dele af (forkert) særskrevne ord hver for sig, ja ofte ligefrem tilskynde brugeren til at bryde komposita i to udelukkende for at "tilfredsstille stavekontrollen". Her vil OrdRet benytte sin viden om komposita, samt morfologisk interdependente træk i nabo-ord, til at markere særskrivningsfejl (fx 'banegårds center').

#### Fejldatabaser

Som tidligere nævnt, udmærker OrdRet sig ved at supplere korrekt-listen med en data-dreven fejl-liste<sup>3</sup>, der genkender fejlstavede ord direkte og sammenholder dem med de forskellige ord, fejlen er blevet rettet til i database-teksterne. Imidlertid har det været umuligt at finde materiale nok til at dække hele sproget og gennemføre en dækkende fejlstatistik. For at sikre at alle rimelige rettelsesforslag kom med for en given fejl, skulle fejllisten derfor gennemgås

---

<sup>1</sup> Constraint Grammar som datalingvistisk metode blev udviklet ved Helsinki Universitet (Karlsson et.al. 1995), og har været brugt til stavekontrol-programmer for fx svensk (Arppe 2000, Birn 2000) og norsk (Hagen, Lane & Trosterud 2001).

<sup>2</sup> Denne morfologiske analysemaskine benyttes også til at vurdere "danskheden" af de ord, brugeren selv ønsker at indføje i brugerordbogen.

<sup>3</sup> Fejldatabasen bygger på af DVO indsamlede og korrigerede ordblindetekster. Dataindsamling og -format har fra starten været tilrettelagt med henblik på en senere integration i den ordbaserede del af stavekontrollen.

manuelt, især for korte ord, hvor små ændringer skaber potentiel lighed med mange andre ord i sproget. Desuden anvendes der en liste over fejlmønstre i ordstammer med variable dele. Den empiriske fejldatabase indeholder over 13.000 opslag, den manuelle godt 5.000.

### **Ordlighed**

For alle andre end de allerbedste stavere er det et gode, at en stavekontrol ikke bare markerer mulige fejl, men også foreslår rettelser ud fra hvilke ord, der ligner det fejlstavede ord mest. Sammenligningsalgoritmerne skal optimeres, fordi det ikke indenfor rimelig tid er muligt at foretage en dybdegående sammenligning med alle over 1 million ordformer i OrdRets database, og vi har derfor genereret specialleksika, der arbejder med "alternative" ordbilleder: Fonetisk transskription, ordstammer, konsonant- og vokalskeletter m.m. Specielt ordblinde laver mange lydligt baserede fejl, og vi har således udviklet en transskriptionsmaskine, der kan gætte udtalen af et fejlstavet ord og sammenligne det med transskriptionerne af eksisterende ord.

### **Forslagsprioritering**

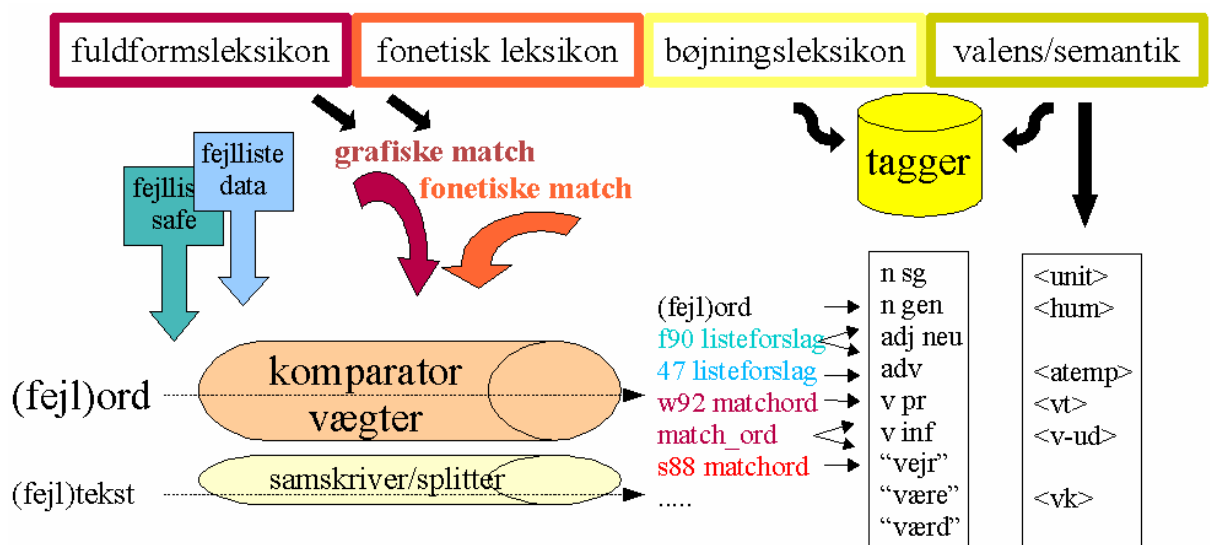
For dårlige stavere, der ikke har en sikker genkendelse af korrektstavede ord heller, er det afgørende ikke kun at finde et antal lignende ord, men også at foretage en god prioritering, optimalt med det rigtige forslag på plads 1 eller 2. OrdRets lighedsvægtning inddrager hele sætningskonteksten til en sådan prioritering (se side 19), fordi selv ord, der ellers ligner meget, kan være meningsløse i konteksten. Lokalt bygger vægtningen på 3 søjler – grafisk lighed, fonetisk lighed og ordets frekvens i normalsproget. Som udgangspunkt for sidstnævnte har vi i projektet kompileret en logaritmisk vægtet frekvensordbog.

### **Fejl i korrektstavede ord**

Et stort problem for en traditionel, listebaseret stavekontrol er at også ellers korrektstavede ord kan være forkerte – enten (1) fordi et fejlstavet ord tilfældigt kommer til at ligne et andet, eksisterende ord (mere sandsynligt, når især ordblinde brugere laver mere end én fejl pr. ord), eller (2) fordi fejlen er rent grammatisk, stilistisk, eller ordstillingsrelateret. OrdRet markerer grammatiske fejl (2) vha. kontekstbaserede regler ("grønne" fejl), mens den håndterer (1) ved at optage også korrektstavede ord i fejllisten som *potentielle* fejl, og råder i øvrigt over et stort homofonleksikon, der kan bruges til at udpege mulige forvekslinger. Imidlertid er der en reel risiko for falsk positive markeringer ved denne metode, og den stiller store krav til OrdRets kontekstuelle disambiguering, som gennemgås i næste afsnit.

## Grammatiske, kontekstbaserede resurser

Kongstanken i OrdRets modulære arkitektur har været at benytte og tilpasse en eksisterende dansk Constraint Grammar parser (DanGram) for kontekstuel at prioritere (vælge/forkaste) rettelsesforslag. På niveau 1 bliver alle ord i en sætning, *plus* rettelsesforslagene, fodret ind i DanGrams morfologiske tagger, der opmærker ord med mulige læsninger som fx *navneord i flertal nominativ*, eller *udsagnsord i datid*. Ord med rettelsesforslag får tildelt summen af alle læsninger for både ordet selv og alle dets rettelsesforslag. DanGram anvender herefter en sprogteknologisk grammatik (2.550 regler) til at bedømme, om en given læsning er lovlig i konteksten eller ikke. Fx kan en læsning som udsagnsord favoriseres, hvis der står et navneord i bestemt form til venstre, mens udsagnsord forbydes til højre for kendeord og bindeord<sup>4</sup>. Grammatikken er i projektet blevet udvidet med flere regelsæt (800 regler), der håndterer den øgede flertydighed rettelsesforslagene medfører, samt kan forholde sig til OrdRets egen opmærkning (grafisk og fonetisk lighed, listeprioritering, frekvensklasser m.m.)



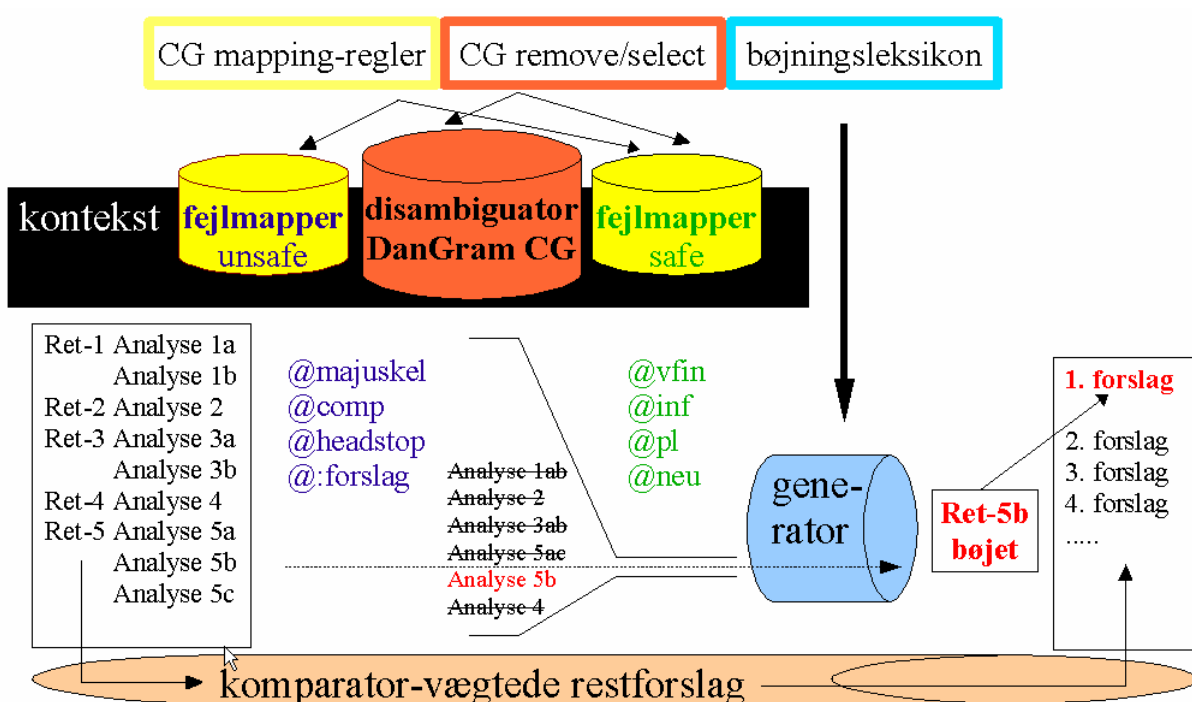
Figur 5: Komparator og tagger

Figur 6 viser samspillet mellem på den ene side den leksikondrevne "komparator", der finder fejl og vægter rettelsesforslag vha. leksikalske resurser (fuldformsleksikon, fonetisk leksikon, bøjningsleksikon og fejldatalister), og på den anden side taggeren, der fastslår ords grundformer (fx "være" for "er", "var", "været" etc.), tilføjer morfologiske læsninger (fx n=navneord, sg=singularis).

<sup>4</sup> Eksemplerne er simplificeret og inddrager kun den umiddelbare kontekst (naboord), men formalismen tillader rent faktisk, at en regel henviser globalt til elementer i eller træk ved hele sætningen (fx "første tidsbøjede udsagnsord i sætningen", eller "næstfølgende navneord uden mellemstående sætningsindledere").

Tallene viser de indbyrdes vægtninger (fx w92=best written match, s88=best spoken match). Til brug ved den kontekstuelle disambiguering tilføjes også oplysninger om relationer imellem ord (fx <vt> for 'transitivt udsagnsord') og semantiske klasser (fx <hum> for 'menneskeord' og <atemp> for tidsbiord).

De grammatiske regler kan ikke kun prioritere eksisterende rettelsesforslag, men også selv tilføje ny information, vha. såkaldte mapping-regler (jf. Figur 6). Disse regler kan fx "mappe" kontekstafhængige rettelsesforslag (@:forslag), samskrivninger (@comp) eller identificere umarkerede sætningsgrænser (manglende punktum: @headstop).



Figur 6: Mapping-regler i OrdRet

På niveau 2 forsøger OrdRets Constraint Grammar-modul at finde grammatiske fejl (grøn i illustrationen), der mappes også på ellers korrektstavede ord. Blandt de vigtigste genkendte fejltypen er kongruensfejl (tal og køn), samt forvekslingen mellem udsagnsords grundform ('-e', @inf) og den nutidsbøjede form ('-er', @vfin). Nedenstående vises en række eksempler for mapping af @-fejltypen, samt en kort omskrivning af de anvendte kontekstbetingelser (kursiv).

Hun har en opfattelse af at **kvinde** (@pl) er bedre til det **merster** (R:meste).  
*(ingen indefinitte entals-substantiver uden prænominaler, - undtagen <mass>)*

Han kan ikke **hører** (@inf) dig.  
*(hjelpeverbumskontekst)*

Han ønsker ikke **og** (@:at) forstyrre.  
*(verbum med infinitivvalens til venstre, infinitiv til højre)*

Hun besøgte **barndoms** (@comp-) veninden.  
*(ubestemt substantiv i genitiv ental før et substantiv i bestemt form)*

Glasset var **fuld** (@sc-neu).  
*(kongruenskrav mellem subjekt og subjektsprædikat)*

Jeg er **træt** (@headstop) Det har **vært** (R:været) en lang dag.  
*(Majuskel og/eller syntaktiske tegn på starten af en ny sætning)*  
*(tillægsmåde 'været' vinder over navneordet 'vært' pga. hjelpeverbum tv.)*

I alt genkender OrdRet over 40 forskellige grammatiske fejltyper. Fordi konsortiet vurderede, at målgruppen ikke ville kunne omsætte denne slags abstrakt fejltypemarkering<sup>5</sup> til egentlige rettelser, blev der udviklet en såkaldt morfologisk generator, der konstruerer de nye, korrigerede bøjningsformer, og indsætter dem på forslagslisten i stil med de forslag, der stammer fra komparatorens leksikalske sammenligninger.

## Evaluering af OrdRets sproglige resurser

I løbet af projektet opstod der et ønske om en formel, kvantitativ evaluering, i tillæg til den løbende, kvalitative evaluering, dels for at kunne kvalitetsbedømme og sammenligne produktet, dels for at kunne træffe funderede beslutninger mht. visse systematiske indstillinger (søgningsdybde, antal forslag, hastighed). Vi besluttede at gennemføre evalueringen på tilfældige uddrag fra konsortiets ordblindetekster, i form af en sammenligning mellem den håndkorrigerede guldstandard (forklaring) og live-kørsler af OrdRet på de samme data. Efter en kontrol af falsk positive fejlmeldinger for de valgte testdata (200 tekster, 32.512 ord) blev der imidlertid fundet yderligere 10-15 % "ægte" fejl, der ikke havde været med i den oprindelige guld-standard, og revisionen er ikke i skrivende stund helt afsluttet.

---

<sup>5</sup> Imidlertid har fejltypemarkeringen en pædagogisk værdi, der måske vil blive udnyttet i en fremtidig version af programmet, fordi den tillader at implementere et læringsaspekt, med eksempler, øvelsesforslag og individuelle fejltypestatistikker.

I forbindelse med evalueringen blev OrdRet kørt uden sin statistiske fejldatabase<sup>6</sup>, men med de manuelt kompilerede fejlmønstre. For at kunne bedømme kvaliteten i fejlforslagsprioriteringen, brugte vi såkaldte vægtningspoint: 1 point, hvis det korrekte forslag var prioriteret højest, ½ point, hvis det stod på plads 2, 1/3 for 3. plads osv. Kun de første 5 forslag blev evalueret. Som almindelig i evaluering af sprogteknologiske programmer, beregnede vi *recall* og *precision*. Førstnævnte fortæller hvor stor en procentdel af alle fejl blev fundet, mens sidstnævnte måler, hvor stor en procentdel af OrdRets fejlmarkeringer var korrekte. Med andre ord er *recall* et mål for falsk negative, mens *precision* bedømmer falsk positive. De to parametre kan slås sammen i en såkaldt *F-Score*, en slags matematisk gennemsnit der dog straffer systemer, hvor en forbedring af den ene parameter går for meget ud over den anden. Med vores vægtningsmetrik skelner vi desuden mellem *simple recall* (om fejlen er markeret overhovedet) og *vægtet recall* (hvor godt det korrekte forslag blev prioriteret). *Vægtet precision* og *vægtet F-Score*, omend lidt uortodoks, blev beregnet tilsvarende.

Til sammenligning blev der også gennemført en manuel evaluering af MS Word for de samme data, og med den samme metrik. Stavekontrollen i MS Word kan som den mest udbredte betragtes som en slags standard, og resultaterne for Word vil derfor i det følgende blive omtalt som *base line*.

	Simple recall	simple precision	simple F-Score	weighted recall	weighted precision	weighted F-Score
all levels	<b>67,9</b>	91,7	78	<b>43</b>	58,1	<b>49,4</b>
safe mode (no green)	54,6	<b>99,1</b>	70,4	34	<b>61,8</b>	43,9
word level (i.e no CG)	59.6	89.8	71.6	32.32	48.7	38.9
word level (no green)	49.1	93.4	64.4	25.2	47.8	33.0
MS Word	53.5	97.3	69.1	19.7	35.7	25.4

Tabel 2: Resultater af kvalitetstest

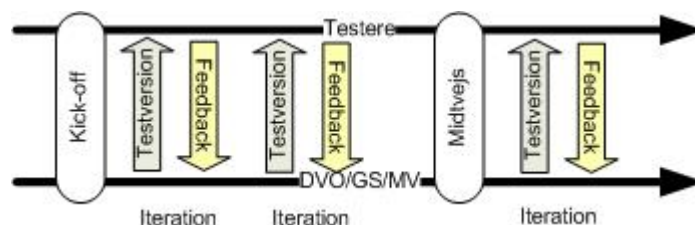
<sup>6</sup> I modsat fald ville evalueringen føre til urealistisk gode resultater, idet testteksterne indgik i det materiale der lå til grund for fejldatabasen.

De anførte tal viser, at OrdRet er klart bedre end en konventionel stavekontrol til at finde fejl og, ikke mindst, til at vægte rettelserforslag i ordblindetekster (vægtet recall 43.0 sammenlignet med Words baseline af 19.7). Prisen, en suboptimal *uvægtet precision*, kompenseres ved at skelne mellem sikre (rød-markerede) og usikre (grøn-markerede) fejl. Uden vægtning ligger såvel recall-gevinsten som precision-tab inden for det "grønne" område, mens "røde" fejl har en uvægtet recall og precision i nærheden af Words baseline (hhv. 99.1 og 54.6). *Med* vægtning for forslagsprioritering ligger samtlige tal, både rød og grøn for både recall og precision, over base line-værdierne (mellem 60% og 115% forbedring). Selvom allerede den kontekst-frie ordniveau-del af systemet er en forbedring mht. forslagsprioritering (vægtet F-Score 33 mod base line 25.4), er det på dette område at Constraint Grammar modulerne har deres hovedvirkning (vægtet F-Score 49.4).

## Ekstern afprøvning af OrdRet

Dette afsnit beskriver den proces, der blev anvendt til afprøvning af OrdRet i forbindelse med inddragelse af eksterne testere. Processen omkring den interne test af programmets sproglige resurser er beskrevet i foregående afsnit.

For at sikre at produktet har den tilsigtede funktionalitet i en virkelig brugssituation, har der været tilknyttet eksterne testere til projektet. De eksterne testere er blevet udvalgt, så de repræsenterer et bredt udsnit af OrdRet's potentielle brugere. Der har således været repræsentanter for ordblinde børn, unge og voksne. Der er tale om speciallærere med it-erfaring, der underviser hhv. ordblinde børn, unge og voksne. Der var tre repræsentanter involveret i projektet og disse repræsentanter afprøvede OrdRet i deres respektive skoler/institutioner. Det betød, at OrdRet blev afprøvet i en brugergruppe bestående af 33 personer.



Figur 7: Testforløb

Selve testen er foregået som illustreret i Figur 7. Hele forløbet blev startet med et kick-off-møde, hvor testernes rolle blev gennemgået og en demonstration af brugergrænsefladen gennemført.

Formålet med mødet var at igangsætte processen og at få tidlig feedback endnu inden, programmeringen var påbegyndt. Herefter blev med planlagte mellemrum udsendt testversioner af programmet til de eksterne testere. Gennem fejlhåndteringsværktøjet Bugzilla og gennem møder og personlig kontakt, har de eksterne testere givet feedback til projektets tre parter.

Feedbacken er herefter blevet opsamlet i en database, og udtræk herfra er anvendt ved de styregruppemøder, hvor næste testversion er blevet planlagt. Nogle af de indkomne rapporter og kommentarer er således blevet indarbejdet i OrdRet, mens andre er blevet afvist eller udskudt til senere versioner af produktet.

På foranledning af Sygekassernes Helsefond indledte vi i foråret 2004 et samarbejde med Havredal Gl. Skole i forbindelse med afprøvningen. To speciallærere skulle deltage som testere. Da OrdRet i foråret 2005 var klar til afprøvning fik Havredal Gl. Skole tilsendt et udførligt instruktionsbrev med en nøje beskrivelse af testprocedure. Havredal Gl. Skole fik ligesom de eksterne testere med planlagte mellemrum udsendt en testversion af programmet. Idet de eksterne testere fandt fejlhåndteringsværktøjet meget kompleks, vurderede vi, at lærerne fra Havredal Gl. Skole ikke skulle involveres i dette, men blot rapportere fejl og forbedringer over mail.

Imidlertid kom samarbejdet ikke til at fungere efter hensigten, idet testerne fra Havredal Gl. Skole midt i afprøvningsperioden meldte ud, at de alligevel ikke havde mulighed for at finde egnede deltagere til at afprøve programmet.

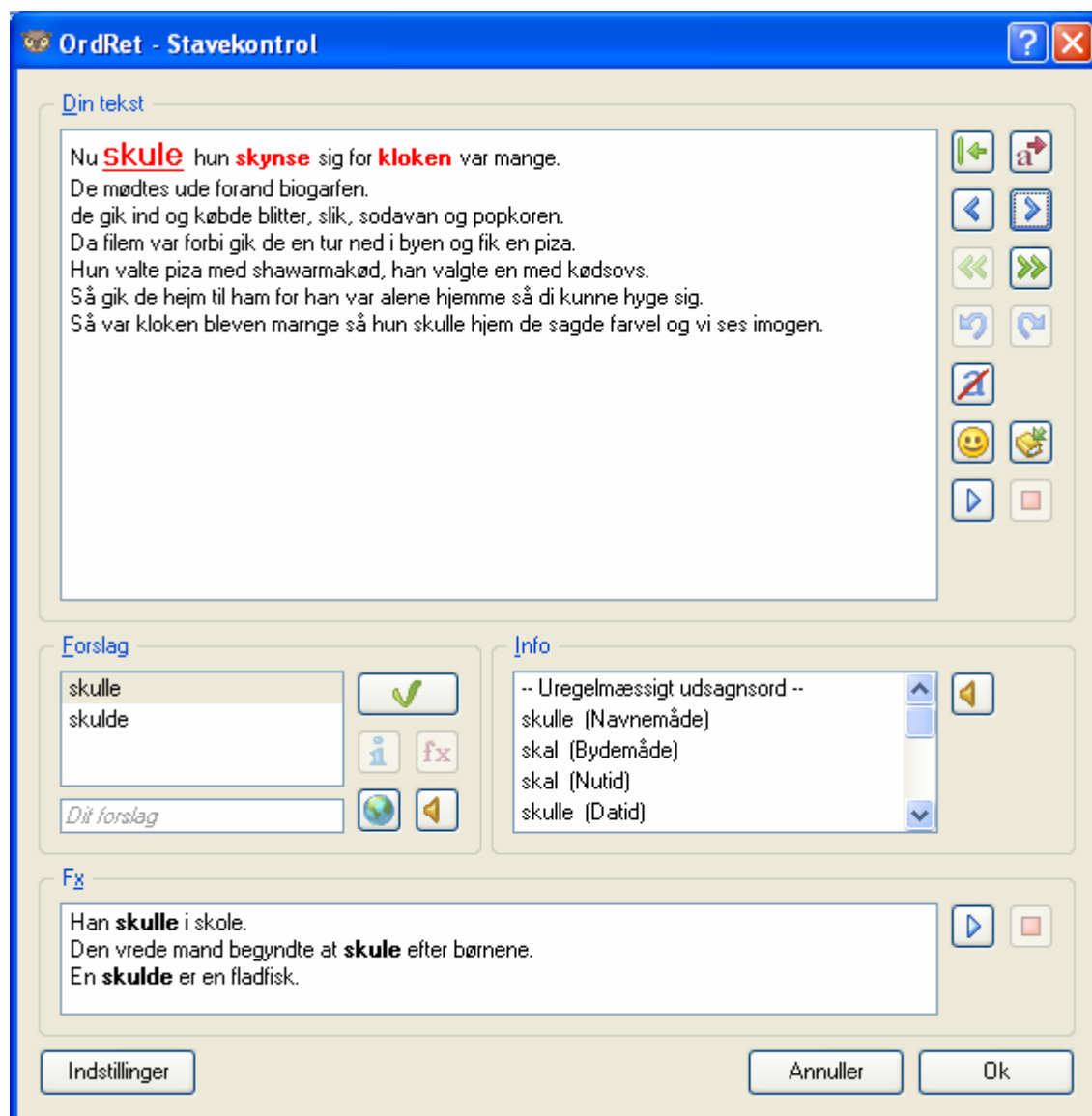
## Præsentation af programmet

I dette afsnit vil vi ganske kort præsentere programmet OrdRet. OrdRet kan det, almindelige stavekontroller kan, men dens styrke er, at den udover at analysere de enkelte ord, forkaster og prioriterer ordforslag ud fra den grammatisk kontekst, som fejlene forekommer i. OrdRet analyserer ikke kun det enkelte fejlord, men den analyserer også hele den sætning, som fejlordet forekommer i. Det betyder, at OrdRet finder mange fejl og de ordforslag, som OrdRet giver, bliver af meget høj kvalitet.

I OrdRet kan man også få hjælp til at skelne mellem ord, der lyder ens eller næsten ens. Når ord, der lyder ens eller næsten ens kommer som ordforslag, giver OrdRet mulighed for at skelne mellem dem ved hjælp af eksempelsætninger med ordene. Eksempelsætningerne er korte, og de er formuleret i et klart sprog. OrdRet kan også give informationer om de enkelte ordforslag. Der er informationer om ordklasse, køn og bøjning.



Når OrdRet er installeret på computeren, lægger programmet sig som en værktøjslinje i Microsoft Word. Når man åbner og lukker programmet Word, åbner og lukker man også OrdRet. Når man ønsker en tekst fejllrettet, kan OrdRet gå i gang med at søge efter fejlstavede ord. Når den har fundet nogle fejlstavede ord, kan den læse dem højt og foreslå korrekt stavning.



Figur 8: OrdRets hovedvindue

På ovenstående skærbillede af OrdRet er alle felter synlige, fordi vi har indstillet OrdRet til at vise disse felter. OrdRets hovedvindue kan tilpasses i størrelse. Øverst til venstre i billedet er tekstfeltet med en tekst og med de røde markeringer af de fejlord, som OrdRet har fundet. Man navigerer i feltet ved at bruge knapperne til højre for tekstfeltet.

OrdRet er et fleksibelt program, som brugeren via diverse indstillingsmuligheder let kan tilpasse til sine individuelle behov, og som således er sikret en optimal tilgængelighed for handicappede.

Fejlfindingsstrategien kan tilpasses den enkelte brugers fejlprofil og skriftsproglige niveau i forhold til beherskelse af syntaks og grammatik. Endvidere kan brugeren efter behov til - og fravælge funktionerne "Info" (information om et ords ordklasse, køn og bøjningsmønster) og "Fx" (enslydende og næsten enslydende ord, der kan være svære at skelne for den ordblinde bruger, vises i eksempelsætninger).

OrdRets brugergrænseflade kan sættes op, så den imødekommer den enkelte brugers behov og ønsker i forhold til skrifttype, -størrelse, -farve og baggrundsfarve.

Mange ordblinde har, udover retskrivningsvanskeligheder, også læsevanskeligheder. Derfor kan alt i OrdRet - brugergrænseflade, hjælp, ordforslag, eksempelsætninger og information om ord - læses op med syntetisk tale. OrdRet bliver lavet i to versioner – en med integreret syntetisk tale og en uden syntetisk tale. Brugere, der ikke i forvejen har en syntetisk tale installeret på deres computer kan anvende OrdRet med integreret talesyntese, mens brugere der i forvejen har en syntetisk tale på deres computere kan anvende denne. For begge versioner gælder det, at OrdRet kan anvendes med digital tale, hvis brugeren har en sådan på computeren. Med de fleksible oplæsningsmuligheder er det sikret, at svage læsere kan få den optimale hjælp til at få rettet en tekst.

OrdRet har en omfattende hjælpefunktion. Overalt i OrdRet kan brugeren få hjælp ved at hente spørgsmålstegnet ned fra hovedvinduetts venstre hjørne og føre det hen over et givent emne eller funktion. Denne hjælp er en "her-og-nu"-hjælp, der forklarer brugeren, hvad emnet eller funktionen kan bruges til. Hjælpefunktionen leder således på rette vej, så brugeren undgår at søge efter hjælp i manualen midt i retteprocessen. Hjælpeteksterne er formuleret i et klart og entydigt sprog.

En mere omfattende hjælp end den, der gives i hjælpefunktionen, findes i OrdRets manual. I OrdRets manual gennemgås alle OrdRets funktioner og muligheder grundigt. Alle funktioner er illustreret med skærmdumps og billeder af knapper og ikoner. Teksterne i manualen er formuleret i et utvetydigt og klart sprog, der imødekommer en målgruppe med læsevanskeligheder. Manualen kan selvfølgelig læses op, hvis brugeren har en syntetisk tale installeret på sin computer.

Brugeren kan gå direkte fra programmets ordforslag til en internetbrowser. På den måde er brugeren sikker på, at søgeordet er korrekt stavet og brugeren kommer direkte til relevante informationer.

## Opsamling og evaluering af projektet

Projektet har henover de sidste fem år, men særligt siden 2003 udviklet den it-baserede stave- og skrivehjælp OrdRet. Med hjælp af gode planlægningsværktøjer, en intensiv møderække og et ansvarligt projektejskab hele vejen rundt, er vi kommet helskindet igennem udførelsesperioden med de leverancer og aktiviteter, der fulgte med, så vi nåede deadline med et absolut tilfredsstillende resultat.

Projektets formål er blevet opfyldt; der er blevet udviklet en stavekontrol, der giver brugeren sproglig støtte og virker som et funktionelt redskab for ordblinde børn, unge og voksne og andre med retskrivningsvanskeligheder.

Afprøvningen af OrdRet blandt kompetente konsulenter og ordblinde unge og voksne viser, at OrdRet er et produkt med store anvendelsesmuligheder.

Kvalitetstestningen af OrdRet's stavekontrol sammenlignet med andre viser, at OrdRet finder mange af de stavfejl, som ordblinde laver og i de tilfælde, hvor OrdRet har flere ordforslag til et fejlord, ligger det intenderede ord meget højt på listen af ordforslag.

At OrdRet har en så stærk fejlfinding skyldes i høj grad, at programmet dels er baseret på en flersidig syntaktisk og grammatisk analyse af den tekst, som brugeren har skrevet og dels, at OrdRet er baseret på en stor database med indsamlede stavfejl fra ordblinde børn, unge og voksne.

## Forventninger til kommercielle resultater af projektet

I løbet af sommeren 2005 forventer projektgruppen, at alle fejl og mangler er rettet. På dette tidspunkt vil OrdRet som kommercielt produkt være klart til salg. OrdRet vil blive produceret og solgt gennem Mikro Værkstedet. Det er igennem kontrakten mellem projektpartnerne aftalt, at hvis der bliver en indtjening, skal den fordeles på følgende måde:

45 % til Mikro Værkstedet, 27 % til GrammarSoft ApS og 18 % til Dansk Videnscenter for Ordblindhed (anvendes til forskning i ordblindes stavfejl). De resterende 10 % skal anvendes til fremtidig vedligeholdelse og udvikling af produktet.

## Perspektiver

Vi håber naturligvis, at OrdRet kan blive den trædesten, der giver afsæt for forbedrede versioner. Som det fremgår af denne rapport, så har vi fået udviklet en rigtig god stavekontrol. Men vi tror også, at OrdRet har potentiale til at udvikle sig til et endnu bedre program. Vi vil her nævne nogle af de udviklingsmuligheder, vi ser i OrdRet.

- Udbygning af hjemmesiden med dialogplatform med brugerne
- Pædagogisk videreudvikling fx med fejltipeinformation til brugerne
- Udvikling af et modul til kontrol af skrivestil
- Udvikling af et tegnsætningsmodul
- Udbygning af fejl databasen
- Effektundersøgelse med OrdRet

Stavekontrollen er et slags historisk mindesmærke, stavefejlenes historie set fra en særlig synsvinkel. Og de vildveje sproget har taget kan jo blive lige så oplysende som det korrekte. Så udover hvad der er fremlagt her i rapporten, er der stadig mange muligheder i datamaterialet for at belyse sammenhænge mellem stavefærdigheder blandt ordblinde børn, unge og voksne.

Men vi modtager gerne kommentarer og forslag til opfølgende analyser og videreudviklinger på hjemmesiden: [www.ordret.com](http://www.ordret.com)

## Referencer

- Arppe, Antti (2000). Developing a grammar checker for Swedish. I: Nordgård, T. (ed.) *Nodalida '99 Proceedings from the 12th Nordiske datalingvistikkdager*, Department of Linguistics, University of Trondheim, p. 13-27.
- Bick, Eckhard (2001). En Constraint Grammar Parser for Dansk. I: Widell, Peter & Kunøe, Mette (eds.), *8. Møde om Udforskningen af Dansk Sprog, 12.-13. oktober 2000*, Århus University, p. 40-50
- Birn, Jussi (2000). Detecting grammar errors with Lingsoft's Swedish grammar checker. I: Nordgård, T. (ed.) *Nodalida '99 Proceedings from the 12th Nordiske datalingvistikkdager*, Department of Linguistics, University of Trondheim, p. 28-40.
- Bruck, M. (1988). The word recognition and spelling of dyslexic children. *Reading Research Quarterly*, 23(1), s. 51-69.
- Hagen, Kristin & Lane, Pia. & Trosterud, Trond (2001). En grammatikkontrol for bokmål. I: Kjell Ivar Vannebo & Helge Sandøy (eds.): *Språkknyt 3-2001*.
- Haven, D. & Nielsen, H. (2004). *Ordblindhed – rapport over spørgeskemaundersøgelsen: "Kommunale tilbud til ordblinde elever"*. Rapport. Dansk Videnscenter for Ordblindhed.
- Karlsson, Fred & Voutilainen, Atro & Heikkilä, Jukka & A. Anttila. 1995. *Constraint Grammar. A Language-Independent System for Parsing Unrestricted Text*. Mouton de Gruyter, Berlin.
- Lennox, C. & Siegel, L. S. (1998). Phonological and ortographic processes in good and poor spellers. I: C. Hulme & R. M. Joshi (red.), *Reading and spelling: Development and Disorders* (s. 395-404). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.
- Lingdys: <http://www.lingit.no/sitefolder/lingweb/>
- Moats, L. C. (1995). *Spelling. Development, disability and instruction*. Baltimore: York Press.
- Moats, L. C. (1996). Phonological spelling errors in the writing of dyslexic adolescents. *Reading and Writing*, 8, s. 105-119.
- Nelson, H. E. (1980). Analysis of spelling errors in normal and dyslexic children I: U. Frith (red.), *Cognitive processes in spelling* (s. 475-493). London: Academic Press.
- Sawyer, D. J., Wade, S. & Kim, J. K. (1999). Spelling errors as a window on variations in phonological deficits among students with dyslexia. *Annals of Dyslexia*, 49, s. 137-159.
- Stava Rätt: [http://www.oribi.se/stava\\_ratt.htm](http://www.oribi.se/stava_ratt.htm)

# OrdRet

Denne rapport redegør for udviklingen af den it-baserede stavekontrol OrdRet.

Man kan læse om projektets livsforløb, udviklingen af programmet og de overvejelser, der ligger bag de sproglige resurser i stavekontrollen. Der er oplysninger om, hvordan programmet fungerer og med hvilke resultater.

Endelig kan man få et indblik i, hvad vi har af ideer til en videreudvikling af programmet.

OrdRet-konsortiet består af:



**Mikro Værkstedet A/S**



**GrammarSoft Aps**



**Dansk Videnscenter for Ordblindhed**

ISBN: 87-990274-1-0