

OrdRet - en ny dansk stavekontrol

Av Julie Kock Clausen

Dansk Videnscenter for Ordblindhed har siden 2000 været ansvarlig for projektet »OrdRet – en ny dansk stavekontrol«.
Hensigten med stavekontrollen OrdRet er, at alle med behov for hjælp til stavning skal have glæde af programmet.
Herudover er OrdRet en stavekontrol, der er udviklet til at være særlig anvendelig for ordblinde.

I 2000 påbegyndte Dansk Videnscenter for Ordblindhed et projektsamarbejde med Mikro Værkstedet, der udvikler it-produkter til undervisningsområdet, og GrammarSoft, der bl.a. med udgangspunkt i undervisningsprogrammet VISL (Visual Interactive Syntax Learning) udvikler it til sprogundervisning. Udviklingen af stavekontrollen er hovedsageligt finansieret af eksterne midler fra It- og Telestyrelsen, herunder Kompetencecenter It for Alle (KIA), Undervisningsministeriets Tips- og Lotto-midler, Egmontfonden og Sygekassernes Helsefond. Tilsammen har projektet fået bevilget et beløb på cirka 3.7 millioner kroner.

Inspirationen til udviklingen af en ny dansk stavekontrol kom primært i 1999, da Dansk Videnscenter for Ordblindhed deltog i en stor nordisk dysleksikonference. Her blev den svenske stavekontrol *Stava Rätt* fremvist. Stava Rätt er en stavekontrol, der er udviklet særligt til ordblinde og er baseret på en omfattende indsam-

ling af voksne ordblindes stavefejl. Stava Rätt var allerede dengang et imponerende program og har siden fået stor anerkendelse og er blevet meget udbredt i Sverige. I en undersøgelse, hvor Stava Rätt sammenlignes med stavekontrollen i tekstbehandlingsprogrammet MS Word viste det sig, at den svenske stavekontrol identificerer og retter 80% af stavefejlene skrevet i en tekst af en ordblind, mens Words stavekontrol kun formår at identificere og rette ca. 30% (Andersson, 2000). Tilsvarende har det norske projekt *Norsk tekstbehandling for dyslektikere – Lingvistisk basert staveverktøy*, forkortet *LingDys*, dokumenteret gode resultater både med hensyn til genfindning og præcision¹ (Stensby m.fl., 2000). I vore nordiske nabolande findes altså allerede stavekontroller, der imødekommer de begrænsninger, ordblinde og andre med retstavningsvanskeligheder møder.

Dansk Videnscenter for Ordblindhed tog initiativ til at nedsætte en projektgruppe med det formål at udvikle en dansk stavekontrol, der var mindst lige så god – og måske endnu bedre – end de nordiske. I projektets levetid har målgruppen udvidet sig fra kun at rumme ordblinde til også at rumme andre mennesker med behov for hjælp til stavning. Ordblindes stavevanskeligheder er i særdeleshed sejlivede og langvarige at få bugt med, men der er tale om en langt bredere gruppe af mennesker, der kan have behov for stavehjælp. På baggrund af de seneste års undersøgelser er et realistisk bud, at ca. 12% af befolkningen har så store problemer med retstavning, at de har behov for mere end, hvad traditionelle stavekontroller kan tilbyde.

Hertil kommer tosprogede, hvis behov for stavehjælp ofte ligeledes rækker ud over almindelige stavekontroller. Det er således projektgruppens mål, at OrdRet bliver en stavekontrol, som rigtig mange vil få glæde af.

Eksisterende stavekontroller

I Dansk Videnscenter for Ordblindhed startede projektforløbet med at skaffe et overblik over de allerede eksisterende stavekontroller og deres hensigtsmæssighed i forhold til ordblinde. I Danmark findes der ikke en stavekontrol, som er særligt udviklet til at imødekomme de stavefejl, ordblinde laver. Langt de fleste mennesker – med og uden retstavningsvanskeligheder – anvender stavekontrollen i Word. Derfor vil stavekontrollen i Word her tjene som eksempel på de vanskeligheder, der ofte ses i forhold til eksisterende stavekontroller. Gennem projektforløbet er udviklingen af Words stavekontrol fulgt nøje, og det må siges, at der gennem årene er sket væsentlige forbedringer, der både imødekommer ordblindes skriftsproglige vanskeligheder og mere generelle retstavningsvanskeligheder. Alligevel kommer stavekontrollen i Word gang på gang til kort. Her nævnes nogle af de mest betydningsfulde mangler.

- Stavekontrollen har ofte ikke et kvalificeret bud på et ordforslag til et givet fejlstavet ord. Stavekontrollen er således ikke god nok til at »gætte« brugers intenderede ord. Og når Stavekontrollen tilbyder ordforslag, er disse ikke hensigtsmæssigt prioriteret. Dette betyder, at der ikke er sikkerhed for, at det intenderede ord kommer som et af de første forslag, men måske først er forslag nummer fem i rækken af mulige ord. Konsekvensen er, at brugeren stilles over for en lang liste af valgmuligheder i identifikationen af den korrekte form 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 den til tider uoverskuelige liste af ord er blandt andet, at Word ikke prioriterer sine ordforslag i forhold til den kontekst, hvori den givne fejl er begået. Dette medfører, at der ofte gives irrelevante alterna-

tiver til det intenderede ord, det vil sige, at der tilbydes ordforslag, der ikke passer ind i den syntaktiske kontekst.

- Stavekontrollen markerer rigtige navneord og specielle fagudtryk som fejl. Den kender fx ikke forholdsvis almindelige sammensatte 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, er den eneste mulighed for at få undersøgt, om formodningen er korrekt, at dele ordet og se efter, om stavekontrollen accepterer de to eller flere enkeltdele, som ordet består af. Dette er dog en meget usikker strategi at forlade sig på, og det ansporer til fejlagtige særskrivninger af ord, der skal sammenskrives.
- Stavekontrollen kan ikke identificere og skelne mellem fejlstavede ord, som er homofone (enslydende) eller næsten homofone. Både homofone og næsten homofone ord forveksles ofte, men fejlmarkeres ikke, fordi der er tale om korrekt stavede ord. Et eksempel på sådanne let forvekslige ord er *bager - bæger - bærer - bære*. Konsekvensen af, at stavekontrollen ikke fejlmarkerer ord af denne type, er, at en sætning som *Du kan købe kager og brød hos en bære* vil gå glat igennem uden fejlmarkering. Årsagen til denne fejl er igen, at stavekontrollen ikke tager konteksten i betragtning, det vil sige, at der ikke tages hensyn til den syntaktiske sammenhæng, som det aktuelle ord indgår i.
- Stavekontrollen kan ikke foreslå korrekt stavning ved stavefejl, der ligger langt fra den intenderede stavemåde. Denne begrænsning medvirker til at indskrænke anvendeligheden af stavekontrollen for ordblinde, fordi de ofte begår stavefejl, som både grafisk, men måske også lydligt ligger langt fra det intenderede ord. Et eksempel er fejlordet *pejget* for ordet *pakket* i sætningen *Far har pejget bilen i flere timer*.
- I Words stavekontrol er der ikke mulighed for oplæsning af ordforslag, menuer, tips og hjælpetekster med syntetisk eller digital tale. Dette er problematisk i forhold til gruppen af ordblinde, der ofte, udover stavevanskeligheder, også har læsevanskeligheder.

De væsentlige mangler, som her er skitseret, har udviklingen af OrdRet taget udgangspunkt i, idet det er hensigten, at OrdRet skal kunne imødekomme de behov, som ordblinde og andre med behov for stavehjælp har, når der skal skrives og staves korrekt.

Indsamling af ordblindes stavfejl

I Danmark er der ikke siden 1965 (Tordrup, 1965) gennemført en større systematisk indsamling og analyse af ordblindes stavfejl. Nutidig viden om danske ordblindes stavning er derfor begrænset og præget af mere eller mindre usystematiske læsepædagogiske erfaringer. Dansk Videnscenter for Ordblindhed gik derfor i 2001 i gang med en omfattende indsamling af tekster skrevet af ordblinde uden brug af nogen form for hjælpemidler. Formålet med indsamlingen var at få en bred viden om ordblindes stavning til brug for den nye danske stavekontrol; derfor blev der indsamlet tekster med stavfejl fra ordblinde børn, unge og voksne, og hver deltagers køn og alder blev påført teksterne.

I alt blev der taget kontakt til 36 amtslige og kommunale institutioner for ordblinde i Danmark. Der blev indsamlet to typer af tekster; frie skriftlige formuleringer og diktater. 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 de ordblindes lærere blev dog opfordret til at foreslå et eller flere emner, så deltagere, der ikke selv havde noget på hjerte, kunne få inspiration til at gå i gang med den skriftlige produktion. Ligeledes blev deltagerne opfordret til at skrive løs uden hensyntagen til eventuelle fejl. Frie skriftlige formuleringer blev medtaget, fordi det er i forbindelse med produktion af denne form for skriftligt materiale, ordblinde vil komme til at anvende stavekontrol. Ligeledes er fri skriftlig formulering en »naturlig« måde at producere noget skriftligt på.

Diktater blev medtaget af flere grunde. Diktater indeholder ofte ord, som er vanskelige at stave, og som derfor kan fremprovokere diverse stavfejl. Endvidere er det forventeligt, at deltagerne i de frie skriftlige formuleringer udeluk-

kende vil benytte sig af ord, som de ikke har 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 mængden af forskellige ord således udvides, og der ville komme flere fejl-stavede ord i det samlede materiale.

Resultater

Vi modtog i alt 1244 tekster, heraf 646 frie skriftlige formuleringer og 598 diktater. For overblikkets skyld blev deltagerne inddelt i tre grupper ud fra alder. I børnegruppen varierede alderen fra 7-12 år, og her deltog i alt 105 børn. I ungegruppen mellem 13-17 år deltog i alt 673 unge, og i voksengruppen mellem 18-62 år deltog i alt 313 voksne. Vi har ikke mange børn med i undersøgelsen. Årsagen til, at der ikke er yngre børn med, er, at stave- og skriveintensiteten i førskolealderen er minimal, og det har derfor ikke været muligt at få deltagende førskolebørn. Årsagen til, at der er 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 er 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 der fra et givet ord altid kan findes tilbage til den tekstsammenhæng, ordet 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 og rigtigt stavede ord identificeret og noteret i databasen. Både stavfejl og rigtigt stavede ord er noteret for at få mulighed for at se på, hvilke ord ordblinde har svært ved at stave og hvilke ord, ordblind rent faktisk kan stave. Dette er en ganske værdifuld information, der kan bidrage med indsigt i, hvilke ord der volder ordblinde særlige vanskeligheder.

Til at identificere stavfejl i frie skriftlige formuleringer er der udviklet et sæt udførlige retningslinjer for rettelser, så måden hvorpå

fejl blev identificeret kunne gøres konsekvent for alle 1244 tekster. I nogle tilfælde har det ikke med tilstrækkelig stor sandsynlighed været muligt at identificere det intenderede ord. I disse tilfælde er ordet registeret som uforståeligt. 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 er der 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 inddeling af fejltyper blev der udviklet en fejltypologi, og ud fra denne er alle stavfejlne analyseret. Fejltypologien er udviklet primært med udgangspunkt i viden om det danske skriftsystem, men også med inspiration fra de ganske få danske og engelske undersøgelser, der findes om ordblindes stavning (Thordrup, 1965; Nelson 1980; Bruck 1988; Moats 1995; 1996; Lennox & Siegel, 1998; Sawyer m.fl. 1999). Der er opstillet i alt 14 fejltyper.

Til selve analysen af stavfejlne blev der udviklet et sæt meget detaljerede retningslinjer. Det skal bemærkes, at hovedparten af de fejlstavede ord indeholder flere fejltyper. Dette skyldes dels, at der kun i en lille del af de fejlstavede ord er begået en enkelt fejl, og dels at alle fejlstavede ord vurderes i forhold til, om de er lydbevarende stavet eller ej. Udover at alle stavfejl er inddelt efter fejltipe, er alle stavfejl inddelt efter ordklasse.

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 et audiologopædisk speciale af Trine Nobelius og Julie Kock Clausen og forhåbentligt vil resultaterne af denne analyse kunne medvirke til at kvalificere OrdRets anvendelse for ordblinde yderligere. Specialet forventes færdigt primo 2006.

De sproglige resurser i OrdRet

Udviklingen af de øvrige sproglige resurser i

OrdRet er blevet varetaget af GrammarSoft. Med udgangspunkt i en Constraint Grammar²-baseret parser for dansk (Bick, 2001) og tilhørende leksikografiske resurser er der skabt et program, der både indeholder traditionelle stavkontrolkomponenter, lydskrivning og kontekst-drevne grammatikker til prioritering af rettel-sesforslag, grammatisk fejlanalyse og generering af bøjningsformer. Arbejdet har været forskningsbaseret og har involveret bl.a. leksi-kografi, udvikling af systemer til sammenlig-ning af ord og anden programmering, korpus-lingvistik, statistisk evaluering samt udarbej-delse og testning af Constraint Grammar-regler.

Listekontrol og sammensatte ord

Rygraden i de fleste traditionelle stavkontrol-programmer består af en liste med ordformer, der er accepterede i normsproget. En traditionel stavkontrol vil ganske enkelt definere en fejl som et ord, der ikke figurerer på listen. Et problem er, at denne metode for dansk kan med-føre mange falsk positive fejlmeldinger, fordi sproget forholdsvis frit sammensætter nye ord. OrdRet har derfor dels inddraget et tekstkorpora til at opbygge et leksikon med sjældne sam-mensatte ord, dels har OrdRet indbygget en morfologisk analysemaskine, der kan genkende ukendte ord som mulige sammensatte ord snarere end som reelle fejl.

Et andet problem med en ren listekontrol er, at listen vil acceptere for eksempel de to dele af fejlagtigt særskrevne ord hver for sig. Som før beskrevet vil stavkontrollen ligefrem tilskynde brugeren til at dele sammensatte ord i to ude-lukkende for at »tilfredsstille« stavkontrollen (fx *banegårds center*). Dette problem vil OrdRet imødekomme ved at benytte sin viden om sam-mensatte ord samt morfologisk uafhængige sammenligninger med naboord til at markere særskrivningsfejl.

Fejldatabaser

OrdRet udmærker sig ved at supplere sin liste af korrekt stavede ord med Dansk Videnscenter for Ordblindheds fejldatabase med ordblindes stavfejl. Med denne datadrevne fejlliste over ordblindes stavfejl, genkendes fejlstavede ord

direkte og sammenholdes med de forskellige ord, fejlen er blevet rettet til i de tekster, der ligger til grund for databasen. Imidlertid er det indsamlede materiale ikke stort nok til at dække hele det danske sprog, og en dækkende fejlstatistik kan således ikke gennemføres på baggrund af de indsamlede stavfejl fra ordblinde. For at sikre, at alle rimelige rettelserforslag for en given fejl kommer med, er fejllisten derfor blevet gennemgået manuelt, især for korte ord, hvor små ændringer skaber potentiel lighed med mange andre ord i sproget. Herudover anvender OrdRet en liste over fejlmønstre i ordstammer med variable dele. Fx stammen *fasciner* med alle dens mulige bøjningsformer og sammensætninger. De variable stammer er en meget effektiv måde at fange endnu ikke sette fejl på. Den empiriske feildatabase indeholder over 13.000 opslag, den manuelle godt 5.000 opslag.

Ordlighed

For svage stavere er det en fordel, at en stavekontrol ikke bare markerer mulige fejl, men også foreslår rettelser ud fra hvilke ord, der ligner det fejlstavede ord mest. Derfor sammenligner OrdRet fejlord med et specialleksikon, der arbejder med »alternative« ordbilleder. De alternative ordbilleder er fx fonetisk transskription af ordet, ordstammer, konsonant- og vokalskeletter m.m. Specielt ordblinde laver mange lydligt baserede fejl, og der er således til OrdRet udviklet en transskriptionsmaskine, der kan gætte udtalen af et fejlstavet ord og sammenligne det med transskriptionerne af eksisterende ord.

Forslagsprioritering

For svage stavere, der ikke har en sikker genkendelse af korrekt stavede ord, er det afgørende ikke kun at finde et antal lignende ord, men også at der fra stavekontrollens side er foretaget en hensigtsmæssig prioritering af disse ord, optimalt med det rigtige forslag på en første- eller andenplads. OrdRets vægtning af ordforslag inddrager hele sætningskonteksten til en sådan prioritering, fordi selv ord, der ellers ligner meget, kan være meningsløse i konteksten. Vægtningen bygger på tre søjler – grafisk lighed, fo-

netisk lighed og ordets frekvens. Som udgangspunkt for sidstnævnte er der til OrdRet udarbejdet en logaritmisk vægtet frekvensordbog.

Fejl i korrektstavede ord

Et stort problem for en traditionel, listebaseret stavekontrol er, at korrektstavede ord kan være forkerte i den kontekst, som de indgår i. Et klassisk eksempel er udsagnsord i grundform skrevet for udsagnsord i nutidsform i sætningen *Her lære vi grammatik*. En traditionel stavekontrol vil ikke opdage en sådan fejl, fordi ordet *lære* er et korrekt stavet ord. OrdRet markerer sådanne grammatiske fejl ved hjælp af kontekstbaserede regler, mens den håndterer ord, der ligner et andet ord ved at optage korrektstavede ord som *potentielle* fejl.

Grammatiske, kontekstbaserede resurser

Kongstanken i OrdRets arkitektur har været at benytte og tilpasse en eksisterende dansk Constraint Grammar parser for ud fra konteksten at prioritere rettelserforslag, det vil sige at vælge eller forkaste rettelserforslag.

På første niveau bliver alle ord i en sætning plus rettelserforslagene sat ind i OrdRets morfologiske analysemaskine, der opmærker ord med mulige læsninger som fx *navneord i flertal* eller *udsagnsord i datid*. OrdRet anvender herefter en sprogteknologisk grammatik, der i alt består af 2.550 regler, til at bedømme, om en given læsning er »lovlig« i konteksten eller ej. Fx kan en læsning som *udsagnsord* favoriseres, hvis der står et navneord i bestemt form til venstre, mens en læsning som *udsagnsord* forbydes til højre for kendeord og bindeord³. Grammatikken er i OrdRet blevet udvidet med endnu 800 regelsæt, der håndterer den øgede flertydighed, som rettelserforslagene naturligt medfører, og som samtidig kan forholde sig til OrdRets egen opmærkning af ord (grafisk og fonetisk lighed, listeprioritering, frekvensklasser med mere).

På andet niveau forsøger OrdRets Constraint Grammar-modul at finde grammatiske fejl, der mappes på ellers korrekt stavede ord. Blandt de vigtigste genkendte fejltypen er kongruensfejl (tal og køn) samt forveksling mellem udsagnsords grundform og nutidsform.

I alt genkender OrdRet over 40 forskellige grammatiske fejltyper. Projektgruppen vurderede, at ordblindemålgruppen ikke ville kunne omsætte denne abstrakte fejltypemarkering til egentlige rettelser. Derfor 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 OrdRets leksikalske sammenligninger. Imidlertid har fejltypemarkeringen en pædagogisk værdi, der måske vil blive udnyttet i en fremtidig version af OrdRet, fordi den tillader at implementere et læringsaspekt med eksempler, øvelsesforslag og individuelle fejltypestatistikker.

Præsentation af stavekontrollen OrdRet

OrdRet kan anvendes i Microsoft Word eller Outlook. Når OrdRet er installeret på computeren, lægger programmet sig som en værktøjslinje i Word eller Outlook. Se Figur 1. Når en tekst skal rettes af OrdRet indsættes den automatisk i programmet.

OrdRet er en stavekontrol, der kan alt det, almindelige stavekontroller kan. Derudover er dens helt klare styrke, at den udover at analysere de enkelte ord, forkaster og prioriterer ordforslag ud fra den grammatiske kontekst, fejlene forekommer i. OrdRet analyserer således ikke kun det enkelte fejlord, men også hele den sætning, fejlordet forekommer i. Det betyder, at OrdRet finder mange fejl og de ordforslag, OrdRet giver, bliver af meget høj kvalitet. Se Forslag-feltet i Figur 2.

Grundpillen i OrdRet er naturligvis den meget stærke stavekontrol. Men i OrdRet er der også medtænkt en pædagogisk dimension. I OrdRet kan brugeren få hjælp til at skelne mellem homofone ord eller næsten homofone ord. Når brugeren skriver ord med homofone alternativer, giver OrdRet mulighed for at skelne mellem dem ved hjælp af eksempelsætninger med de homofone alternativer. OrdRet har i alt ca. 8000 eksempelsætninger, som alle er korte og formuleret i et entydigt, klart og præcist sprog. Se Fx-feltet i Figur 2.

OrdRet kan også give grammatiske informati-

oner om de enkelte ordforslag. Der er informationer om ordklasse, køn og bøjning. Dette kan være en hjælp for brugere med en vis grammatisk viden, når der skal skelnes mellem forskellige ordformer. Se Info-feltet i Figur 2.

På Figur 2 ses et skærmbillede af OrdRets hovedvindue. På billedet er alle felter synlige, fordi programmet her er indstillet til at vise disse felter. Men brugeren kan selv sammensætte sit hovedvindue efter behov – man kan fx fravælge at få vist eksempelsætninger (Fx-feltet) og information om ord (Info-feltet). OrdRets hovedvindue kan desuden tilpasses i størrelse, og skriftstørrelse, -type og -farve kan sammen med fontfarve indstilles efter ønske. OrdRet er således et fleksibelt program, som brugeren via diverse indstillingsmuligheder let kan tilpasse sine individuelle behov, og som således er sikret en optimal tilgængelighed for alle.

Øverst i OrdRets hovedvindue ses tekstfeltet med en tekst og med de røde og grønne markeringer af de fejlord, som OrdRet har fundet. De rød-markerede ord er fejlord, som OrdRet vurderer som *reelle* fejl, mens de grøn-markerede ord er *potentielle* fejl. Herudover kan OrdRet indstilles til at blå-markere homofone ord. Brugeren navigerer frem og tilbage mellem de enkelte ord eller hele sætninger og annullerer rettelser i feltet ved at bruge knapperne til højre for tekstfeltet.

I OrdRets *Indstillinger* kan programmets fejlfindingsstrategi tilpasses den enkelte brugers fejlprofil og skriftsproglige niveau i forhold til beherskelse af syntaks og grammatik. Endnu en finesse er, at 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.

Mange ordblinde har, udover retstavningsvanskeligheder, også læsevanskeligheder. Derfor kan alt i OrdRet - brugergrænseflade, hjælp, ordforslag, eksempelsætninger og information om ord - læses op med syntetisk tale. OrdRet udkommer i to versioner – en med integreret syntetisk tale og en uden syntetisk tale. Brugere-

re, 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 computer direkte 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.

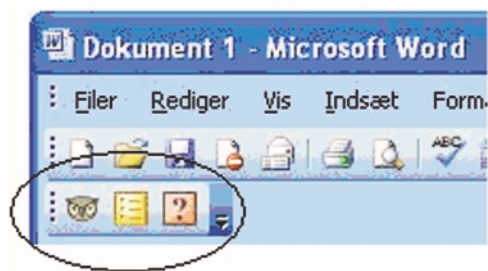
OrdRet har en omfattende hjælpefunktion, der indeholder hjælpetekster, der er formuleret i et klart og entydigt sprog. Overalt i OrdRet kan brugeren få hjælp ved at hente et spørgsmålstegn ned fra hovedvinduet venstre hjørne og føre det hen over et givent emne eller funktion. Denne hjælp er en »her-og-nu«-hjælp, 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.

En mere omfattende hjælp end den, der gives i hjælpefunktionen, findes i OrdRets manual. I OrdRets manual gennemgås alle OrdRets funk-

tioner og muligheder grundigt. Alle funktioner er illustreret med skærmdumps og billeder af knapper og ikoner. Teksterne i manualen er ligeledes 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.

Intern evaluering af OrdRets sproglige resurser

I løbet af projektet opstod der et ønske om en formel, kvantitativ evaluering af OrdRets sproglige resurser, dels for at kunne kvalitetsbedømme og sammenligne programmet med andre stavekontroller, dels for at kunne træffe hensigtsmæssige beslutninger med hensyn til OrdRets indstillinger (søgningsdybde, antal forslag, hastighed og så videre). Det blev besluttet at gennemføre evalueringen på tilfældige uddrag af de allerede indsamlede ordblindetekster i form af en sammenligning mellem de rettelser, der allerede var foretaget af de indsamlede tekster og live-kørsler af OrdRet på de samme data.



Figur 1: OrdRet i Word.



Figur 2: OrdRets hovedvindue.

	Simple recall	simple precision	simple F-Score	weighted recall	weighted precision	weighted F-Score
OrdRet						
Alle niveauer ⁴	67,9	91,7	78	43	58,1	49,4
OrdRet						
Niveau I	54,6	99,1	70,4	34	61,8	43,9
OrdRet						
Niveau II	59.6	89.8	71.6	32.32	48.7	38.9
OrdRet						
Niveau III	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 1: Resultater af evaluering.

I forbindelse med evalueringen blev OrdRet naturligvis kørt uden sin statistiske fejldatabase, men med de manuelt udarbejdede fejlmønstre. For at kunne bedømme kvaliteten i ordforslagsprioriteringen, blev der brugt såkaldte vægtningspoint: 1 point, hvis det korrekte forslag var prioriteret højest, $\frac{1}{2}$ point, hvis det stod på plads 2, $\frac{1}{3}$ for plads 3 osv. Kun de første 5 forslag blev evalueret. Som det er almindeligt i evaluering af sprogteknologiske programmer, beregnedes *recall* og *precision* (genfindning og præcision). Førstnævnte faktor måler hvor stor en procentdel af alle fejl, der blev fundet, mens sidstnævnte måler, hvor stor en procentdel af OrdRets fejlmarkeringer, der 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.

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

De anførte tal i tabel 1 viser, at OrdRet er klart

bedre end en konventionel stavekontrol til at finde fejl og - ikke mindst - til at vægte rettel-sesforslag i ordblindetekster (vægtet recall er 43.0 sammenlignet med Words baseline på 19.7). Prisen, en mindre optimal *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 området for »grønne« fejl, 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). Vi kan således konkludere, at OrdRet er væsentligt bedre end stavekontrollen Word til både at identificere fejlord og komme med de korrekte ordforslag.

Ekstern afprøvning af OrdRet

For at sikre at OrdRet har den tilsigtede funktionalitet i en virkelig brugssituation, har der været tilknyttet eksterne afprøvere til projektet. De eksterne afprøvere er blevet udvalgt, så de repræsenterer et bredt udsnit af OrdRets potentielle ordblinde brugere. I spidsen for afprøvnin-gen har der stået tre speciallærere med massiv it-erfaring, der underviser henholdsvis ordblinde børn, unge og voksne. De tre speciallærere afprøvede OrdRet på deres elever i deres re-

spektive institutioner. Det betød, at OrdRet blev afprøvet i en brugergruppe bestående af i alt 33 personer.

Selve testforløbet blev startet med et informationsmøde, hvor afprøvernes rolle blev gennemgået og en demonstration af OrdRets brugergrænseflade gennemført. Formålet med mødet var at igangsætte processen og at få tidlig feedback endnu inden, at selve programmeringen af OrdRet var påbegyndt. Herefter blev OrdRet afprøvet med ordblinde brugere, og der blev med planlagte mellemrum udsendt testversioner af programmet til de eksterne afprøvere. Gennem et internt it-baseret fejlhåndteringsværktøj og gennem møder og personlig kontakt, har de eksterne afprøvere givet særdeles brugbar feedback til projektets tre parter.

Feedbacken er herefter blevet opsamlet i en database, og udtræk herfra er anvendt i planlægningen og udviklingen af de løbende testversioner. Nogle af de indkomne kommentarer er således blevet indarbejdet i OrdRet, mens andre er blevet afvist eller udskudt til senere versioner af programmet.

Perspektiver for projekt OrdRet

Som det fremgår af denne artikel, er OrdRet blevet en rigtig god og gedigen stavekontrol. Men OrdRet har potentiale til at udvikle sig til et endnu bedre program. Udviklingsmuligheder for OrdRet kunne være:

- Pædagogisk videreudvikling fx med fejltypen information til brugerne
- Udvikling af et modul til kontrol af skrivestil
- Udvikling af et modul til kontrol af tegnsætning
- Udbygning af fejldatabasen med ordblindes stavefejl, evt. i form af særligt tilrettelagte dikterer, så den i højere grad dækker det danske sprog

Effektundersøgelse af OrdRet

Kommentarer til OrdRet og forslag til opfølgende analyser og videreudviklinger modtages meget gerne på OrdRets hjemmeside www.ordret.com

Hvornår lanceres OrdRet?

OrdRet vil blive produceret og solgt gennem Mikro Værkstedet. OrdRet forventes klar til lancering på den internationale læsedag torsdag den 8. september 2005.

Denne artikel er baseret på *Rapport om projekt OrdRet – en it-baseret stavekontrol*. Mikro Værkstedet A/S, Grammarsoft Aps & Dansk Videnscenter for Ordblindhed, 2005.

Rapporten kan i sin helhed hentes fra www.ordret.com

Noter:

- 1) Førstnævnte faktor måler hvor stor en procentdel af alle fejl, der blev fundet, mens sidstnævnte måler, hvor stor en procentdel af fejlmarkeringerne, der var korrekte[0]. Med andre ord er *genfindning* et mål for falsk negative, mens *præcision* bedømmer falsk positive.
- 2) Constraint Grammar som datalingsvistisk metode er for omfattende at forklare i nærværende artikel. Metoden blev udviklet ved Helsinki Universitet (Karlssoon et.al., 1995) og har været brugt til stavekontrolprogrammer for fx svensk (Arppe 2000: Birn 2000) og norsk (Hagen, Lane & Trosterud, 2001). For yderligere information se http://beta.visl.sdu.dk/constraint_grammar.html.
- 3) Eksemplerne er simplificerede 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 ud-sagnsord i sætningen« eller »næstfølgende navneord uden mellemstående sætningsindledere«).
- 4) Niveauerne refererer til forskellige indstillingsniveauer internt i OrdRet.

Julie Kock Clausen,
projektmedarbejder,
Dansk Videnscenter for Ordblindhed

Referencer:

- Andersson, Bodil (2000). Stava Rätt - ett stavningsprogram som tänker språkligt. *Dyslexi - aktuellt om läs- och skrivsvårigheter*, p. 14-17.
- Arppe, Antti (2000). Developing a grammar checker for Swedish. I: Nordgård, T. (ed.) *Nodalida '99 Proceedings from the 12th Nordiske datalingsvistikkdager*, 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 datalingsvistikkdager*, Department of Linguistics, University of Trondheim, p. 28-40.
- Bruck, Magareth (1988). The word recognition and spelling of dyslexic children. *Reading Research Quarterly*, 23(1), p. 51-69.

- Hagen, Kristin & Lane, Pia. & Trosterud, Trond (2001). En grammatikkontrol for bokmål. I: Kjell Ivar Vannebo & Helge Sandøy (eds.): *Språknytt 3-2001*.
- Haven, Dorthé. & Nielsen, Helen (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, Carolyn & Siegel, Linda. 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, Louisa C. (1995). *Spelling. Development, disability and instruction*. Baltimore: York Press.
- Moats, Louisa C. (1996). Phonological spelling errors in the writing of dyslexic adolescents. *Reading and Writing*, 8, s. 105-119.
- Nelson, Hazel E. (1980). Analysis of spelling errors in normal and dyslexic children I: Uta Frith (ed.), *Cognitive processes in spelling* (s. 475-493). London: Academic Press.
- Sawyer, Diane. 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.ori.se/stava_ratt.htm
- Stensby, Torbjørn m.fl. (2000). Norsk tekstbehandling for dyslektikere. Lingvistisk basert staveverktøy (ling-Dys). Sintef tele og data, rapport STF40 A00054.
- Tordrup, S. A. (1965). Stavefejl og fejltyper hos elever fra femte normalklasse og fra femte og sjette læseklasse. *Skolepsykologi*.

Links:

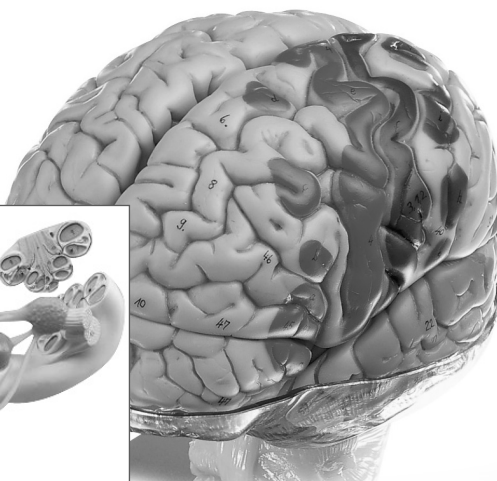
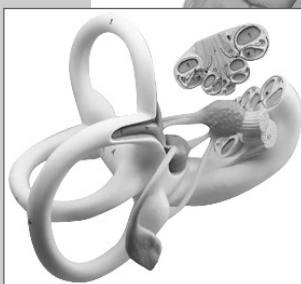
- Dansk Videnscenter for Ordblindhed: www.dvo.dk
- GrammarSoft: www.grammarsoft.com
- Mikro Værkstedet: www.mikrov.dk
- OrdRet: www.ordret.com

Undervisningsmateriale

til Audiologopæder f.eks.

- Øremodeller
- Larynxmodeller
- Hjernemodeller
- Plancher m.m

Rekvirér kataloger



Kvalitetsmodeller fra SOMSO



SCANTERAPI

Fr. Bajersgade 26
DK-8700 Horsens
Tlf. 70 255 299 • Fax. 70 255 219