

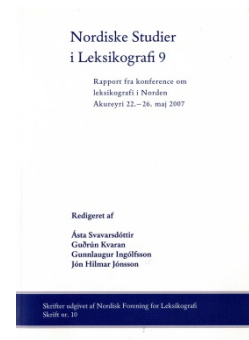
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Jysk Ordbog – Rapport fra en digital ordbog

Forfatter: Inger Schoonderbeek Hansen

Kilde: Nordiska Studier i Lexikografi 9, 2008, s. 209-218
Rapport fra Konference om leksikografi i Norden, Akureyri 22.-26. maj 2007

URL: <http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive>



© Nordisk forening for leksikografi

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Jysk Ordbog – Rapport fra en digital ordbog

The Peter Skautrup Centre has been publishing the Dictionary of Jutland Dialects (Jysk Ordbog) on the internet since October 2000, and everyone with access to a computer connecting the internet has free access to all articles from A to Hæ-. The multilayered site contains links to dialect geographical maps, references to applied literature, a list of editorial abbreviations, an overview of phonetical signs, and furthermore links to information on the editorial background of the dictionary, i.e. geographical distribution and the assembling of the large and varied material. As the process of editing is very complex and, more important, in a way hazardous due to out of date software, plans are being made to update the technical platform of the dictionary. This new, far more coherent platform will consist of first of all an editorial (webbased) database, written in a XML standard. As the Centre has succeeded in finding some of the acquired financing funds, the final phase will start in securing and making more easily available the online dictionary.

1. Forhistorien

11. oktober 2000 blev der på Institut for Jysk Sprog- og Kulturforskning (som Peter Skautrup Centret på daværende tidspunkt hed) holdt en reception i forbindelse med lanceringen af Jysk Ordbog på internettet. Efter mange års forberedelser var det lykkedes, med teknisk hjælp fra IT-virksomheden UNI-C, at offentliggøre færdigredigerede afsnit af dialektordbogen for alle med adgang til computer og internetopkobling under adressen: www.jyskordbog.dk.

I første omgang var det ordbogsafsnittene A–F brugerne kunne søge i, men i dag er afsnittene til og med Hæ- lagt på nettet (se figur 1) (sidste ord er **hævte**, verb. ‘true (med en kæp el. lign.)’).

Hjemmesiden opfylder i store træk redaktionens ønsker om præsentationen af Jysk Ordbog. På hjemmesidens frontside finder man oplysninger om redaktionens adresse, dens sammensætning gennem tiderne og praktiske oplysninger med hensyn til at downloade ordbogens særlige skrift. Linkene på sidens venstre halvdel fører til oplysninger om ordbogens redaktionelle baggrund med information om ordbogens emne (“Hvad dækker ordbogen?”), ordbogens forskellige korpusser (“Ordbogens samlinger”), artiklernes opbygning, ordbogens hjælpemidler (lydskrift, ordbogens kilder, dialektgeografiske kort mv.). Endelig er her også et link til ordbogens historie, der starter tilbage i 1932 da Peter Skautrup oprettede Aarhus Universitets første rigtige institut. Det centrale link på frontsiden fører naturligvis læseren til opslag i ordbogen.

Jysk Ordbog

Hvad dækker ordbogen?
 Område
 Periode
 Ordforråd

Ordbogens samlinger

Artiklernes opbygning
 Ordklasser
 Artiklernes hoved
 - Opslagsform
 - Udtale
 - Bøjning
 Kantet parentes
 Artiklernes krop
 - Betydning
 - Citater

Hjælpemidler
 Lydskrift
 Områder mv.
 Forkortelser mv.
 Kildeangivelser
 Dialektkort

Ordbogens historie

Velkommen til Jysk Ordbog

Slå op i ordbogen her (A - Hæ)

Resten af ordbogen vil blive lagt ud i takt med redaktionen. Har du spørgsmål til afsnit Hø - Å, kan du klikke [her](#), skrive spørgsmålet - og få svar fra os senere.

Ordbogen anvender en særlig lydskrift, som kan downloades [her](#)

- til PC
- til MAC
- til LINUX

Peter Skautrup, Centret for Jysk Dialektforskning
 Aarhus Universitet
 Jens Chr. Skous Vej 4
 Nobelparken, bygn. 1483
 8000 Aarhus C.

Tlf.: 8942 1913
 Fax: 8942 1912
 E-mail: jyskordbog@hum.au.dk

Jysk Ordbog er grundlagt af Peter Skautrup.

Indsamlet og prøveredigeret af Peter Skautrup, K. Flingsgaard, Niels Age Nielsen, Ove Rasmussen og Hans Veirup.

Redigeret af Ove Rasmussen, Viggo Sørensen, Torben Arboe, Inger Schoonderbeek Hansen og Nina K. Grøtthauge.

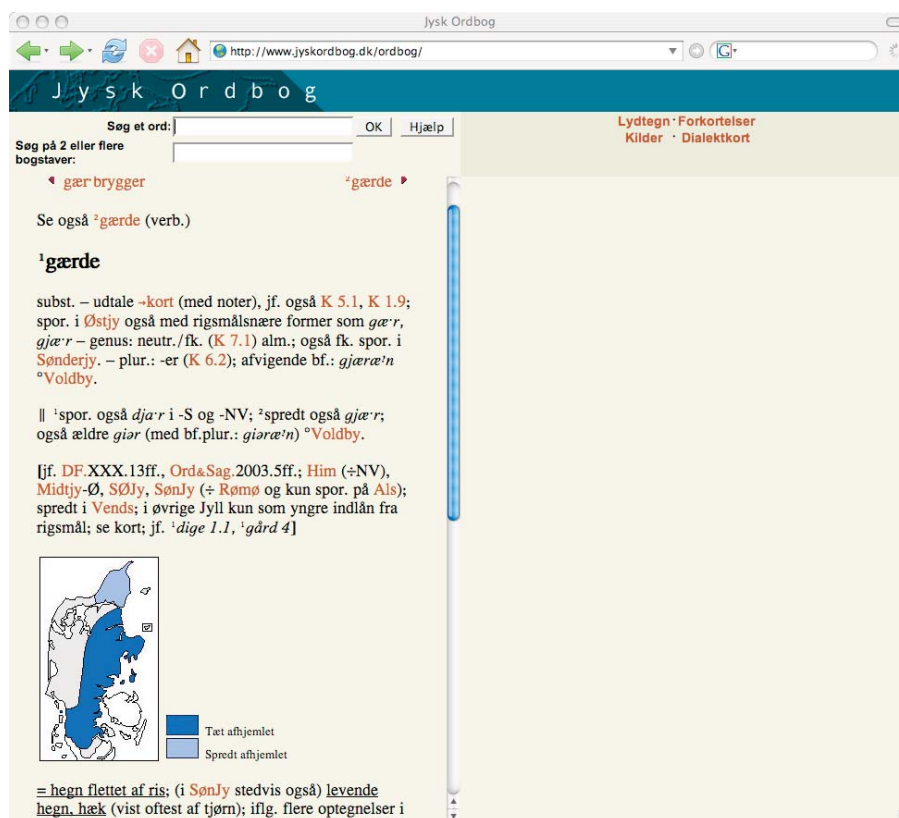
Omsat til skærmudgave af Nina K. Grøtthauge og Kirsten Lyshøj.

Implementeret på Internet af UNI-C ved Bengt Moss-Petersen.

Web-siderne er redigeret af Ove Rasmussen og Viggo Sørensen. Alle henvendelser bedes rettet til [web-editor Kirsten Lyshøj](#).

Figur 1. Jysk Ordbogs frontside

Ved hjælp af to søgefelter på opslagssiden kan man foretage opslag i ordbogen. I øverste søgefelt, "Søg et ord", kan man taste et helt ord ind. Og i søgefeltet nedenunder kan man, som nyt tiltag fra 2003–2004, søge ord i Jysk Ordbog ved at taste 2 eller flere startbogstaver af det ord man leder efter (se øverst i venstre spalte på skærbilledet fra figur 2). Øverst på opslagssidens højre halvdel fører linket "Lydtegn" brugeren hen til informationer om ordbogens lydskrift. Lydskriften er relativt grov, baseret på rigsmålets ortografi og Danias lydskrift. Via de øvrige link øverst på højre side kan brugeren finde informationer om de forkortelser der bruges i ordbogen, de kilder ordbogen anvender, og endelig via linket "Dialektkort" komme frem til en gruppe af dialektkort, vi hyppigt henviser til, vedrørende grovere eller finere inddelte dialektområder, forekomst (eller ej) af stød og længde i de jyske dialekter, jyske modsvarigheder til rigssproglige eller gammeldanske vokaler og konsonanter, særlige jyske bøjningsforhold mv.



Figur 2. Opslag ¹gærde 'hegn flettet af ris ...' i www.jyskordbog.dk

Alle link i ordbogen er markeret med rødt, også i de enkelte artikler (jf. figur 2). Opslag i Jysk Ordbog gør det således muligt at bladre frem og tilbage til foregående og efterfølgende artikel ved at klikke på henvisninger øverst i artiklen (se figur 2: “◀ gær·brygger” og “²gærde ▶”). Samme link er også placeret i bunden af den side artiklen optræder på (ses ikke på figur 2). I de tilfælde et opslag kunne føre til to eller flere artikler i Jysk Ordbog er der via et link mulighed for at komme frem til disse; det kan fx være homonymer (se figur 2: “Se også ²gærde (verb.)”).

Den enkelte artikel er normalt opbygget på følgende måde (se figur 2):

1. Hoved: opslagsform; ordklassebetegnelse; oplysninger om ordets udtale og bøjning (med links til ordbogens dialektkort); evt. noter til udtaleangivelsen (eller til et helsides udtalekort).
2. Etymologisk parentes: evt. etymologi; ordets udbredelse (evt. med henvisning til et udbredelseskort); henvisning til evt. synonymer.
3. Kroppen: ordets oversættelse(r) til rigsdansk; efterfølgende citater, der illustrerer ordets brug.

I den enkelte artikel vil brugeren ofte via link blive ført til:

- a) Udtaleoplysninger, fx ved linket “udtale →kort” til udtalekort i skærmens højre side (evt. med noter); ved henvisning til standardkort i ordbogens atlasdel (dialektkortene) over accentforhold mv. ved link som “(K 5.1)”, “(K 1.9)” etc.
- b) Geografiske oplysninger ved link fra “se kort” til tilføjet udbredelseskort; ved link fra områdebetegnelser som Østjy, RougsøH; ved link fra ordbogens vigtigste kilder (de såkaldte punktoptegninger, fx “°Voldby”) til et kort, der åbnes i et separat vindue.
- c) Kildeoplysninger ved link fra ordbogens litteraturforkortelser til en liste over ordbogens kilder, der åbnes i højre halvdel af siden.

Sådan fungerer Jysk Ordbog i dag, meget kort fortalt.¹

2. Nuværende teknisk platform

Bag Jysk Ordbogs meget brugervenlige grænseflade ligger en noget omstændelig og risikofyldt redaktions- og publikationsproces.

Selve redaktionen foregår ad følgende seks trin:

- I Redaktøren redigerer på baggrund af Jysk Ordbogs seddel- og spørgelistemateriale² sit ordbogsafsnit i et almindeligt tekstbehandlingsprogram (Word 5.1 for MacIntosh!); alle dialektgeografiske kort bliver tegnet i Adobe Illustrator (version 11.0.0).
- II Når et afsnit er færdigredigeret, gennemlæses det af en anden redaktør for til sidst at havne til endelig finpudsning, opsætning og omsætning til Word 97-fil, hvilket foregår på Centrets eneste pc.
- III Word 97-filen konverteres via en specielt designet makro til en ny fil i en slags HTML-format.
- IV Denne (pseudo) HTML-fil flyttes via en browser fra pc'en til en af UNI-C's servere ved hjælp af et filhåndteringssystem (WebDAV).
- V (Pseudo) HTML-en konverteres til en (rigtig) HTML via et af UNI-C specielt designet programmodul, et såkaldt CGI-script.
- VI Ny artikler (eller tilføjelser til eller ændringer i allerede publicerede artikler) kræver at et større omgivende afsnit af ordbogen opdateres sam-

1 Yderligere beskrivelse af www.jyskordbog.dk: Torben Arboe (2003) “Jysk Ordbog på internettet” i *Nordiske Studier i Leksikografi* 6. Rapport fra Konference om leksikografi i Norden. Tórshavn 21.–25. august 2001. s. 31–41

2 Jysk Ordbogs seddelsamling består af mere end 3 millioner alfabetisk og topografisk ordnede papirsedler opbevaret i 14 kartoteksskabe med i alt knap 1300 skuffer. Derudover består spørgelistematerialet, dvs. besvarelser (fra gennemsnitligt 500 meddelere) af ca. 1200 spørgelister (af gennemsnitligt ca. 20 spørgsmål), der for 75%’s vedkommende er it-lagret, af oplysninger der svarer til godt to tredjedel af seddelsamlingen.

let. Ergo: Jo flere artikler der redigeres, jo længere tager det at opdatere ordbogen.

Den ovenfor beskrevne noget komplekse redaktions- og publikationsproces medfører en række risici:

- Ad I: Word 5.1 er et stærkt forældet tekstbehandlingsprogram.
- Ad II: Tyveri eller nedbrud af pc'en er en stadig fare.
- Ad III: Makroen er ensidigt kompatibel med Word 97, men ikke senere udgaver af Word; en autoriseret nyinstallation af den gamle software er ikke mulig da den ikke længere distribueres af Microsoft; det er desuden ikke sikkert at den specielt designede makro overhovedet kan fungere sammen med senere versioner af Word eller senere versioner af styringssystemet Windows.
- Ad V: Til trods for at systemet kun har haft ganske få nedbrud eller driftsproblemer i alle disse år, er det usandsynligt at UNI-C (endsige andre) på ret kort sigt råder over den fornødne ekspertise mht. de enkelte moduler da disse er baseret på 8-10 år gamle web-teknologier og applikationer (hvilket er *meget gammelt* i edb-verdenen).

Strengt taget er rekruttering af ny redaktører også et problem men det er en helt anden historie.

Processen er altså for kompliceret, for usikker, for langsom og gør brug af for gammelt programmel. Processen bør fremover baseres på åbne standardkomponenter, være mindre kompleks, være (mere) fremtidssikret; den bør simpelthen opdateres.

Kort sagt: Publikationen af Jysk Ordbog på nettet kræver en ny teknisk platform!

3. Fremtidig teknisk platform

Da det i sin tid var UNI-C der stod bag implementeringen af Jysk Ordbog på internet, og de lige siden har stået for hosting af Jysk Ordbog på en af deres servere, var det oplagt at UNI-C så det som sin opgave at komme med et forslag til en forbedret platform for www.jyskordbog.dk.³

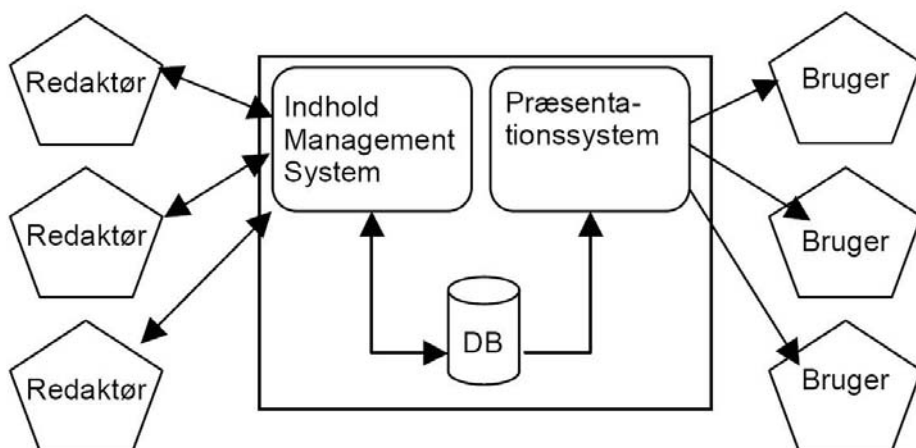
Den ny version af www.jyskordbog.dk tænkes opbygget som et samlet system centralt placeret på en hovedserver. Alt materiale, altså tekst, kort mv., skal ligge centralt på denne server, og der skal derfor ikke flyttes informationer mellem de enkelte elementer af systemet. Selve redaktionen vil foregå online via en browser idet redaktørerne kan arbejde uafhængigt af hinanden.

³ I den forbindelse har UNI-C – efter aftale med Centret – udarbejdet to rapporter: “Forslag vedrørende teknisk platform for www.jyskordbog.dk”, november 2004, og “Tilbud om foranalyse til udvikling af ny platform for www.jyskordbog.dk”, oktober 2006.

I store træk vil den ny version se således ud:

- Det samlede system er mindre kompleks og mere fejlsikker end det gamle idet det bygger på en XML-standard.
- Datagrundlaget bliver mere struktureret, hvilket giver mulighed for senere at eksportere data (i XML-format) til en 3. part, fx via link fra andre netbaserede ordbøger mv.
- Der anvendes et let læseligt programmeringssprog, så systemet lettere kan vedligeholdes og modificeres fremover (evt. også i ny ordbogssammenhænge).

Grafisk fremstiller UNI-C Jysk Ordbogs ny version som det fremgår af figur 3.



Figur 3. Jysk Ordbogs ny teknisk platform efter UNI-C's forslag (jf. note 3)

Det samlede system indeholder en todelt database hvor den ene del indeholder data om ordbogens indhold, management og system mens den anden del af systemet udgør selve præsentationen. Redaktørerne redigerer fx via en browser direkte i systemet (indhold), og ad denne vej kan fx en webdesigner ændre ordbogens management og (til dels) system. Brugerne har naturligvis kun adgang til selve præsentationen af ordbogen.

I forhold til det tekstbehandlingsprogram (Word 5.1) vi bruger i dag, vil det ny, modulært opbyggede redigerings- og publiceringssystem betyde en væsentlig effektivisering af redaktionsprocessen i Jysk Ordbog. Særligt håber vi på at en del af Jysk Ordbogs store og differentierede henvisningsapparat i højere grad bliver automatiseret i den ny teknisk platform. For mange manuelle indtastninger på området medfører i dag en for høj fejlprocentage og en for stor andel i korrekturarbejdet.

4. Jysk Ordbogs henvisningsapparat

Jysk Ordbogs mange forskellige typer henvisninger kan opdeles i *interne* og *eksterne* henvisninger. De interne henvisninger forbinder tekstenheder inden for ordbogen selv, de eksterne henvisninger peger ud til hvad vi kalder ordbogens fællesfaciliteter, dvs. dialektkort, kildefortegnelser mv. I figur 4 ses henvisningernes opdeling i interne og eksterne henvisninger (til dels på basis af artiklen fra figur 2). Der vises om henvisningen p.t. oprettes automatisk eller indtastes af redaktøren, og om de er opmærket eller ej. En firkantet klamme [markerer at henvisningen er forberedt men endnu ikke implementeret, mens en tuborgklamme { markerer at henvisningen endnu ikke er forberedt men ønskes implementeret i fremtiden. De åbne pile i figur 4 viser i hvilken retning vores ønsker med hensyn til den ny platform går, dvs. fra mange manuelle til mange automatiske handlinger.

Interne henvisninger		Eksterne henvisninger	
Manuelt	Automatisk	Manuelt	Automatisk
til opslagsform – indtastet i database med side- og lydformer	til foregående og kommende artikel – i præsentationen markeret med ◀ ▶	til kildefortegnelse – opmærket foran og bag kildeforkortelsen ◊ ⇒	
[til anden artikel – opmærket @ foran fx synonym] ⇒		til atlasdel – opmærket fx “(K 5.1)” ⇒	
		til udtalekort – opmærket “→ kort” ⇒	
		til spaltekort – opmærket “se kort” ⇒	
		{til artikler i andre netordbøger etc.}	

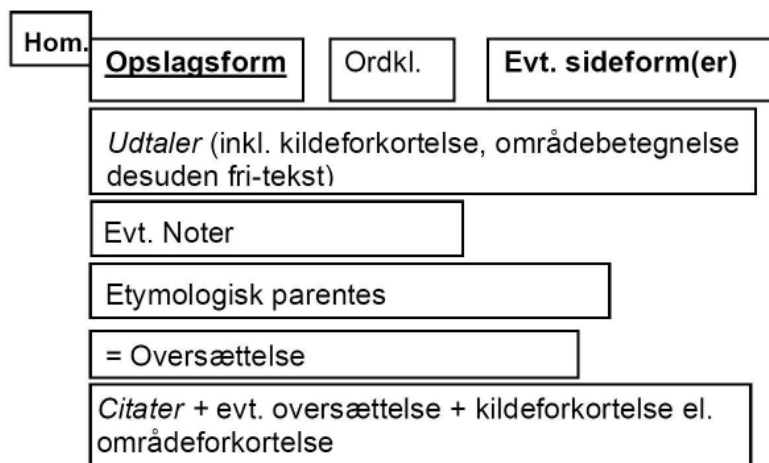
Figur 4. Jysk Ordbogs henvisningsapparat

Den ny platform vil kunne gavne og ikke mindst aflaste redaktionen i forhold til selve indtastningen af dataene, de enkelte artikler. Som sagt bruger vi i dag et almindeligt tekstbehandlingsprogram, og her er det redaktøren der har ansvar for evt. slåfejl, forglemmelser mv. Der bliver naturligvis læst grundig korrektur på hvert enkelt afsnit men der er utrolig mange detaljer at holde rede på: korrekte forkortelser (for områder og kilder), korrekte tegn som punktum, semikolon, komma – for ikke at tale om korrekt skift mellem fonte (fed, kursiv, understreget, antikva, særtegn), så det er svært at undgå småfejl i selve teksten. Dertil kommer opmærkningen til de forskellige eksisterende og planlagte interne og

eksterne link i ordbogen, fx de små hak foran og efter en kildeforkortelse, en bestemt pil foran en henvisningsform, et bestemt mellemrum foran og efter selve opslagsformen etc. Lige nu resulterer fx forveksling af et tegn eller forglemmelse af et punktum i en kildehenvisning i at der ikke oprettes link, og at brugeren altså ikke kan føres videre til kildefortegnelsen. Den ny platform skulle helst automatisk kunne kontrollere korrekt anvendelse af tegn, fonte og forkortelser; på andre punkter må den ikke hindre den enkelte redaktør i at give en artikel en vis form for individuelt, personligt præg).

5. Fremtidsplaner

Hvilke muligheder skal den ny platform til Jysk Ordbog have? Hvis man forestiller sig at databasen i den ny platform består af bokse der skal indeholde bestemt information, skal de enkelte bokse defineres ganske nøjagtigt så fejl eller forglemmelser udbedres automatisk eller eventuel fejlmeldes. For nogle af boksene ville fx gælde at skrifttypen i dem er defineret til at være fed og understreget mens redaktøren i andre bokse har større frihed, fx med hensyn til formuleringen i den etymologiske parentes (se figur 5).



Figur 5. Skitse af Jysk Ordbogs ny redigeringsdatabase

Databasen kunne være opbygget således at der i redaktionen først må tages stilling til om et opslagsord har homonymer, hvilket markeres ved små, hævede tal foran opslagsordet i fed og lidt hævet over linjen (jf. også figur 2). Næste boks indeholder opslagsformen som i præsentationen altid er fed og understreget. Dernæst indtastes ordklasse; boksen kunne fint fungere som rullevindue der indeholder de forskellige muligheder. Dernæst en boks der kan rumme eventuelle sideformer, der altid er fede. Så kommer en boks der skal give lidt større frihed, nemlig til at gøre rede for opslagsformens udtale. Udtalerne er i præsentationen i kursiv mens

al anden tekst er i antikva. De kilde- og områdeforkortelser der tastes ind her, må kunne tjekkes automatisk. Noteapparatet skal i præsentationen indeholde små hævede tal samt selve noteteksten, der refererer til udtaleoversigten eller/og et udtalekort. Noterne indeholder også kilde- og områdeforkortelser – der automatisk skal kunne kontrolleres. Den etymologiske parentes indeholder alle typer henvisninger, der automatisk skal kunne kontrolleres, mens oversættelsen er mere fri, dog optræder (hovedparten af) den altid som understreget tekst i præsentationen. Efterfølgende citater er altid i kursiv mens eventuelle oversættelser er i antikva; også her er der mange område- og kildehenvisninger.

I alle bokse skal der være indbygget en form for henvisning til en anden del af ordbogen. Fx skal homonymnummer i kombination med opslagsformen kunne linkes til artiklens udtalekort; boksen med eventuelle sideformer skal kunne linke til ordbogens søgefelt, så brugeren også kan finde disse afvigende former; område- og kildeforkortelser i boksen med udtaler skal forbindes med oversigten over områder hhv. kildefortegnelsen mv.

Redigeringsværktøjet skal kunne rumme flere forskellige layout så redaktøren fx også kan skrive sin artikel i en slags trappediagram der beskriver ordbogsstrukturen bag hver enkelt artikel. Desuden skal redaktøren frit kunne skifte mellem det ene og det andet layout og mellem de forskellige bokse eller elementer i strukturen, fx sådan at han/hun ikke er bundet til altid at skulle bringe en oversættelse (hvilket vi ikke gør ved ord der kun er afhjemlet fra en enkelt kilde).

Jysk Ordbog har et stærkt presserende behov for ny teknisk platform. I første omgang skal der laves en foranalyse der beskriver hver enkelt komponent i den ny platform, hvad det kræver af hard- og software, hvilke problemer man eventuelt måtte støde på mv. Centret skal her via en workshop bidrage med sit kendskab til ordbogens struktur og sine ønsker med hensyn til præsentationen. En del af dette arbejde er allerede gjort ved etableringen af vores nuværende netudgave mens andet skal revideres da den ny platform anvender ny værktøjer der med stor sikkerhed giver ny muligheder, fx med hensyn til intertekstuelle henvisninger.

Der er også planlagt en workshop om redigeringsværktøjet iLEX (udviklet af Jens Erlandsen), så dets fordele og mangler i forhold til Jysk Ordbog kan afdækkes. iLEX er efterhånden det tool der hyppigst anvendes i ordbogsbranchen i Danmark, fx hos DSL, Gyldendal, og det åbner klart nogle perspektiver. Det vil være en fordel, men ikke en nødvendighed, for det videre forløb hvis eksisterende delkomponenter til en ny platform kunne anvendes.

Endelig er det UNI·C's plan på basis af de to workshoper at udarbejde en rapport som skal indeholde en anbefaling om det videre arbejde med den ny platform. Rapporten vil give et bud på hvilken XML-standard der skal arbejdes med, hvilken editor der skal bruges (altså det redigeringsværktøj redaktørerne skal arbejde med til daglig), og hvordan det videre forløb skal se ud frem til det endelige system.

Siden konferencen i Akureyri maj 2007 har projektet omkring fremtidssikringen af www.jyskordbog.dk været igennem en turbulent periode. Først opsagde UNI-C ensidigt deres samarbejde med os, og hermed faldt alt det planlagte på jorden. Ikke længe efter så fremtiden dog lysere ud da Jysk Ordbog blev medspiller i det langt større projekt DK-CLARIN, hvis mål er lettere digital adgang for såvel forskere som offentligheden til humanistisk og socialvidenskabeligt kildemateriale. Udviklingsarbejdet bliver nu forestået af ekspertise fra Det Humanistiske Fakultets egen IT-afdeling i tæt samarbejde med partnerne i DK-CLARIN-projektet. En stor del af finansieringen til foranalyse og udvikling af en ny platform er dermed kommet på plads, og samtidig vil der blive mulighed for gensidig linking til CLARIN's øvrige partnere.

Inger Schoonderbeek Hansen
f. 1970
Peter Skautrup Centret for Jysk Dialektforskning
Nordisk Institut, Aarhus Universitet
Jens Chr. Skous Vej 4, bygn. 1483, 6. etage
8000 Århus C
jysis@hum.au.dk