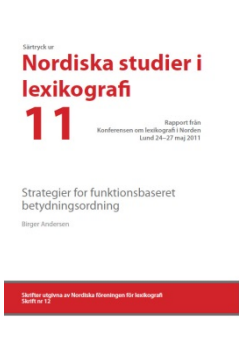


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Flødeboller, mælkebøtter og smørkringler - Om leksikalsk-semantiske problemstillinger i arbejdet med Ømålsordbogen	
Forfatter:	Lena Wienecke Andersen	
Kilde:	Nordiska Studier i Lexikografi 11, 2012, s. 92-102 Rapport från Konferens om lexicografi i Norden, Lund 24.-27. maj 2011	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for lexicografi

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 lexicografi (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.

Flødeboller, mælkebøtter og smørkringler

Om leksikalsk-semantiske problemstillinger i arbejdet med Ømålsordbogen

Lena Wienecke Andersen

In my PhD project, I study how word meanings are ‘captured’ and defined in *Ømålsordbogen*, the Dictionary of the Danish insular dialects, and I study and try to develop new practices. Semantic relations is one of the central points in the project. In this article, I examine two problems concerning the extraction of semantic relations within a non-digitized collection of dialect words (already in or going to be in the dictionary). Since we want to mention semantically related words in the dictionary, we basically need to know what words we have in the collection of paper slips. As the words are historical and most of them unknown, general knowledge and introspection cannot be used. So, first, I will show one way to get an overview of the various words. Second, I outline the possibility to update the work with semantic relations – which so far has been done manually and non-digitized – by means of a Wordnet-inspired tool of modern standard Danish, *DanNet*. But it is argued that the geographical variation in the data has to be taken into consideration; it has to be incorporated that some of the synonyms in the data are actually equivalents. Overall, I argue for an onomasiological approach to the words when working with word meanings, even though the dictionary itself is alphabetically organized.

Nøgleord: DanNet, dialektologi, leksikalsk semantik, onomasiologi, seddelsamling, semantiske relationer, Ømålsordbogen

1. Indledning

Overordnet set kan man skelne mellem to tilgangsvinkler til et ordforråd: Den semasiologiske tilgang hvor ordene er det styrende princip, og den onomasiologiske tilgang hvor begreberne er det styrende (Bergenholtz et al 1997, Svensén 2004, Geeraerts 2010). Med udgangspunkt i mit ph.d.-projekt vil jeg i denne artikel (A) præsentere en fremgangsmåde til at få overblik over et geografisk varieret ordforråd og (B) diskutere hvordan man kan forholde sig til denne vari-

ation i arbejdet med leksikalsk-semantiske relationer. De to problemstillinger illustrerer nødvendigheden af den onomasiologiske tilgang, gerne kombineret med den semasiologiske.

2. Baggrund

Mit ph.d.-projekt der har arbejdstitlen: ”Fra virkelighedens verden til ordbogens – et studie i datastrukturering med henblik på betydningsdefinitioner i ordbøger”, placerer sig i krydsfeltet mellem dialektologi, leksikografi og leksikalsk semantik. Konkret leksikografisk og dialektologisk beskæftiger jeg mig med *Ømålsordbogen*, en dansk dialektordbog der beskriver dialekterne på Fyn, Sjælland, Lolland, Falster og de omliggende øer. Ordbogen er alfabetisk ordnet, og med de 10 bind der foreløbig er kommet, er redaktionen omtrent halvvejs. Konkret leksikalsk-semantisk skal jeg afprøve om strukturen i det digitaliserede begrebsnetværk over moderne standarddansk DanNet kan overføres til *Ømålsordbogens* materiale, og i så fald om DanNet kan bruges som analyseværktøj i forbindelse med de traditionelle dialekter.

3. Problemstilling A: overblik over ordforrådet

Som mange andre ordbøger vil vi i *Ømålsordbogen* gerne angive semantiske relationer, og her støder vi på første meget banale, men alvorligt mente problemstilling: Hvilke ord findes i det materiale vi arbejder med? Og det bringer præsentationens titel på bane:

I moderne standarddansk er en *flødebolle* en spiselig chokoladeovertrukket skumbolle (jævnfør norsk: kokosbolle, svensk: kokosboll), en *mælkebøtte* er en plante med gul blomst der af mange betragtes som ukrudt (norsk: løvetann, svensk: maskros), og en *kringle* er et stykke bagværk med remonce og for eksempel rosiner i (det samme i norsk, *kringla* i svensk).

I de klassiske dialekter på de danske øer var en *flødebolle* en serveringskande til fløde og en *mælkebøtte* en beholder til transport af mælk. *Smørkringlen* vender jeg tilbage til senere.

Mælkebøtte og *flødebolle* er eksempler på at selv om vi på *Ømålsordbogen* arbejder med dansk sprog, eller en del af det, så er der grænser for hvad man som redaktør kan bruge sin almene viden til. Vi sidder med ord fra de danske ødialekter, hovedsageligt fra perioden 1850-1920, og i det materiale findes der mange ord som vi enten slet ikke kender, eller som ikke betyder det samme længere. Redaktørens problem er egentlig ikke betydningsdefinitionen for det pågældende ord; det er faktisk mere reglen end undtagelsen at man skal lave

noget detektivarbejde når man redigerer en dialektordbog. Men hvordan henviser man på kryds og tværs mellem ordbogsartikler, herunder til artikler som endnu ikke er redigeret og måske først bliver det om mange år?

Man bliver nødt til at skaffe sig et overblik over ordforrådet; hvis man for eksempel skal henviser til synonymer, så skal man nemlig vide hvilke synonymer der findes. Vi kan ikke bruge introspektion, og eksisterende synonymordbøger dækker heller ikke. Hvad gør man så? Hvilke ord fandtes i de danske ømål? Hvilke ord forekommer i *Ømålsordbogens* samlinger?

4. Ømålsordbogens seddelsamling

Som i mange andre ordbogsprojekter arbejder vi med sedler. Vi har rundt regnet tre millioner ikke-digitaliserede ordbogssedler. Hvor mange ord, lemmaer, det drejer sig om, er der ingen der ved. Men heldigvis er vores seddelsamling inddelt i emner; og inden for hvert af disse emner står sedlerne så alfabetisk. De cirka 60 emner afspejler den verden som dialekterne eksisterede i, nemlig landbokulturen som den så ud for godt 100 år siden. Der er for eksempel et emne der hedder *gård*, et der hedder *pløjning*, og et der hedder *mælk*. En sådan emneinddeling giver langt bedre mulighed for at få overblik over ordforrådet; inden for hvert enkelt emne vel at mærke.

Oprindeligt var det tanken at alle emner skulle gennemgås systematisk, altså genstand for genstand, proces for proces, før den alfabetiske ordbogsredaktion. Ved en samlet gennemgang af alle ord i emnet fik man dels overblik over ordforrådet, dels over de semantiske relationer heri. Som minimum oplyste man synonymer, men også andre relationer kunne registreres, for eksempel del-helhed og over-underbegreber. Filosofien var at man kun kan beskrive ordforrådet, for eksempel i en ordbog, når man kender genstandene og sammenhængene.

Hvad angår funktionsord, er en almindelig alfabetisk/semasiologisk gennemgang af materialet nok. Men hvad angår alle indholdsord, vil jeg sige at en systematisk/onomasiologisk analyse nærmest er nødvendig – i hvert fald for at få registreret alle de ukendte ord og lumske ord som *flødebolle* og *mælkebøtte* som der gemmer sig mange af i *Ømålsordbogens* materiale.

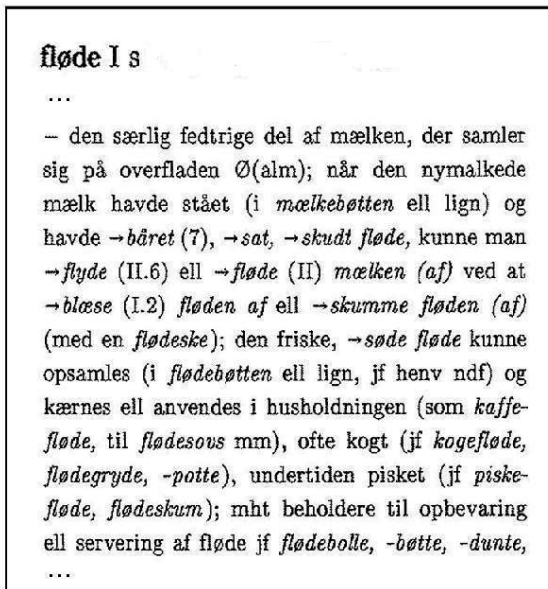
Når sådan et arbejde har fundet sted, er der ideelt set ingen tvivl om hvad der skal henvises til på kryds og tværs i ordbogen. Og det er vældig praktisk når nu ordbogen redigeres og udgives alfabetisk fra en ende af.

I praksis fik man desværre kun gennemgået omkring halvdelen af de cirka 60 emner. Derefter gik den alfabetiske ordbogsredaktion i gang, og man havde ikke resurser til at gøre begge dele.

5. Emnet *mælk*

Et af de emner der ikke nåede at blive gennemgået, er emnet *mælk*, og derfor skal jeg arbejde med det i mit ph.d.-projekt. Det omfatter ord for produkter som mælk, ost og smør, processer som at malke og at kærne, redskaber som en malkespand og en kærne, personer som malkepige og -komtesse og så videre. Kort sagt de ord som de dialekttalende brugte når de talte om emnet. Det drejer sig om i alt cirka 1.200 lemmaer fordelt på 10-13.000 sedler. Mælkeordene er hidtil kun blevet gennemgået alfabetisk, et for et. Det kan forsvares at gennemgå emnet systematisk nu selv om den alfabetiske ordbogsredaktion er lige knap halvvejs, idet de centrale begreber og ord inden for emnet er at finde i den endnu uredigerede del af alfabetet, for eksempel kærne, mælk, ost og smør. Umiddelbart er fløde det eneste centrale begreb der er redigeret.

Et uddrag af flødeartiklen i *Omålsordbogen* ses i figur 1:



Figur 1. Uddrag fra *Omålsordbogen*, bind 4.

I en ”almindelig” ordbog ville man måske have nøjedes med: ”den særlige fedtrige del af mælken der samler sig på overfladen”, men *Omålsordbogen* fortsætter. I artiklen forsøges nævnt alle de ord og begreber som fløde indgår i en relation med.

Meget kan gå galt her. For hvad gør man egentlig når emnet ikke er blevet gennemgået systematisk, og man har over 1.000 lemmaer fordelt på over 10.000 sedler?

Dels gennemser man sine sedler, i det her tilfælde flødesedler, og måske dukker der andre relevante ord og begreber op. Dels kan man gennemse den spørgeliste der har lagt grunden for mælkesamlingen (Møller 1938), og ud fra den kan man få ideer til hvor i alfabetet man kan lede efter relevante ord. I forhold til hvor mange ord der henvises til fra flødeartiklen, kunne jeg forestille mig at redaktøren også har gennemset hele seddelsamlingen af mælkeord. Endelig kan man selvfølgelig studere alle mulige andre kilder.

I mangel af bedre laver vi således en ad hoc-gennemgang af sedler, spørgelister og seddelsamlinger for at få overblik over hvilke relationer det aktuelle ord indgår i. Men det ville naturligvis være at foretrække hvis der forelå en fuldstændig systematisk gennemgang af hele emnet. Og dét er derfor en af opgaverne i mit ph.d.-projekt: At lave en systematisk gennemgang af mælkesamlingen.

6. Overblik over *mælk*

Det første jeg gjorde, var at indtaste alle mælkeordene i en database. Med en elektronisk base (modsat en seddelsamling på papir) kan de samme data sorteres på forskellige måder blot ved at trykke på en knap; når man vel at mærke har indtastet de oplysninger der skal sorteres efter. Som minimum indtastede jeg lemma, ordklasse, en grov semantisk kategorisering og ofte også stikord til en definition. Den grove semantiske kategorisering, som er baseret på intuition, har været nødvendig for mig for at få overblik; så jeg ikke skulle arbejde med 1.200 enheder svarende til de 1.200 lemmaer, men så jeg i stedet havde et overskueligt antal enheder at arbejde med.

I skrivende stund er jeg ved at gennemgå genstanden kærne systematisk. En lang række ord har fået mærkaten ”redskab/kærne”; det er de ord der betegner det redskab man brugte til at fremstille smør med. Muligheden for semantisk sortering ved hjælp af denne kategorisering har gjort at jeg hurtigere har kunnet afgrænse og overskue relevante ord end ellers.

Der er overordnet set to slags kærner: Kærner hvor beholderen står fast, og delene indeni er bevægelige, og kærner hvor beholderen er bevægelig, med faste dele indeni – og inden for disse typer er der igen flere typer og igen flere typer. Der er godt 100 betegnelser i materialet for helheden kærne og alle dens dele hvoraf nogle ses nedenfor:

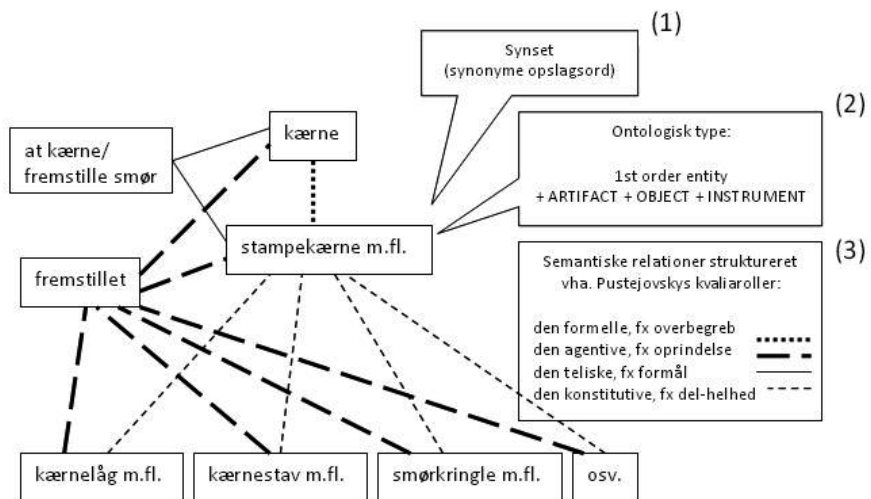
Næste skridt i mit ph.d.-projekt er at afprøve om et begrebsnetværk kan være et nyttigt redskab når man skal arbejde med de semantiske relationer i de danske ømål. Dette arbejde introduceres kort i det følgende.

7. Begrebsnetværk og semantiske relationer

Som nævnt indledningsvis vil jeg afprøve om DanNet kan bruges på mine data. Jeg vil ikke komme nærmere ind på baggrunden for DanNet og dets internationale forløber Wordnet, men blot henvise til DanNets hjemmeside samt Pedersen et al 2010.

DanNet er et digitaliseret begrebsnetværk for moderne standarddansk; et komplekst net der ideelt set skulle bestå af en masse knuder forbundet af en masse linjer; knuderne repræsenterer begreber og betegnelser for dem, linjerne repræsenterer forskellige semantiske relationer.

I figur 3 ses den første brøkdelen af et mælkenetværk, nemlig begrebet stampekerne og dets umiddelbare omgivelser i nettet. I det videre arbejde vil stampekerne skulle kobles til flere ord der igen er koblet til flere ord og så videre.



Figur 3. Stampekærne-synset i et DanNet-lignende netværk.

Udgangspunktet i figuren er begrebsknuden for stampekerne; i samme begrebsknude vil ordene stavkærne og stangkærne og så videre skulle placeres. I Word-

netterminologien er der tale om en gruppe synonymer, et synset, jævnfør (1) i figuren.

Næste skridt er at begrebsknuden skal kategoriseres. Efter et katalog af ontologiske typer skal alle begrebsknuder have en mærkat, jævnfør (2). I det her tilfælde er der tale om en konkret genstand, det vil sige en såkaldt "1st order entity", mere specifikt en menneskeskabt genstand beregnet som redskab. Derefter kobler man andre ord og synsets på ved hjælp af fire kategorier af semantiske relationer, jævnfør (3):

Overbegreb:	kærne
Oprindelse:	menneskeskabt, det vil sige fremstillet
Formål:	at kærne, at fremstille smør
Del-helhed:	stampekærnen består af en række dele i form af synsets for låget, staven, smørkringlen og så videre.

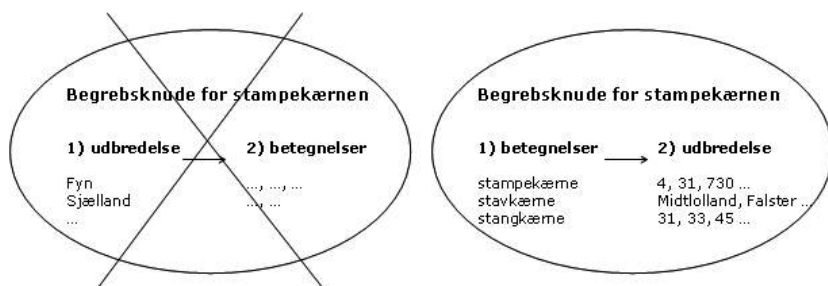
Dertil kommer at stampekærnen overbegreb såvel som dele også er fremstillede. Og ligeledes har overbegrebet kærne samme formål som stampekærnen.

Og sådan skal alle mælkeordene efterhånden placeres og forbindes på kryds og tværs i begrebsnetværket – når problemstillingen som præsenteres i det følgende, er blevet løst.

8. Problemstilling B: den geografiske variation

Betegnelserne for for eksempel *smørkringlen* er ikke bare almindelige synonymmer der ubetinget kan anbringes i ét synset. I *Ømålsordbogen* kan betegnelser for den samme genstand være (1) "almindelige synonymmer" hvor to ord mere eller mindre betyder det samme. Jeg har for eksempel en seddel hvor samme person har angivet at kunne sige både *trille* og *trisse* om træringen. De to betegnelser ville kunne placeres i et synset uden problemer. Men betegnelserne kan også være (2) "geografiske synonymmer", eller ækvivalenter om man vil, der dækker over at man i én dialekt bruger et ord, i en anden dialekt bruger et andet ord for det samme. Denne geografiske synonymi skal inkorporeres i Wordnetmodellens synsets på en eller anden måde hvis den skal kunne bruges i arbejdet med *Ømålsordbogen*.

I figur 4 ses to begrebsknuder hvori der på to forskellige måder er taget højde for den geografiske variation. Det venstre eksempel er krydset over idet det ikke er en mulig løsning på grundlag af *Ømålsordbogens* materiale.



Figur 4. To eksempler på hvordan den geografiske udbredelse kan indarbejdes i en begrebsknode. Eksemplet til venstre er ikke muligt på grundlag af *Ømålsordbogens* materiale.

I begrebsknuden til venstre er tanken at hver dialekts betegnelse(r) holdes samlet sådan at man til slut ville kunne lave et net med alle begrebsknuder i hver enkelt dialekt, for eksempel fynsk. Dette forudsætter dog dels at man helt *principielt* mener at man kan afgrænse dialekterne skarpt; dels at det kan lade sig gøre i *praksis*.

I begrebsknuden til højre ses vidt forskellige angivelser af udbredelse angivet med topografiske numre (Skjelborg 1967) eller områdebetegnelse. Det topografiske nummer 31 omfatter et enkelt sogn, Vejby Sogn, mens det topografiske nummer 4 omfatter hele Nordsjælland. Ligeledes ses også de omfattende lokaliteter Midtjylland og Falster. På en stangkærneseddel ses igen sogn nummer 31. I Vejby kunne man altså både sige stampekerne og stangkærne om denne type kærne; her er tale om almindelig synonymi. I nogle tilfælde er de samme ord således almindelige synonymmer, i andre tilfælde er de samme ord geografiske synonymmer.

I Wordnettankegangen kan der laves et net per (national)sprog; således ville der (vel) principielt også kunne laves et net per dialekt. På grundlag af *Ømålsordbogens* materiale kan en skarp afgrænsning af dialekterne dog ikke lade sig gøre i praksis. På *Ømålsordbogens* sedler er angivet en lokalitet, som regel et sogn (for eksempel Vejby), men undertiden mere upræcist en egn eller et dialektområde (for eksempel Midtjylland, Falster). Det mest præcise vi kan gøre, er derfor at registrere de forskellige typer lokalitetsangivelser sådan som det ses i begrebsknuden til højre, altså angive udbredelsen så fint- eller grovmasket som den nu

engang er registreret.¹ Og derfor vil vi hellere tale om geografiske synonymer end om et antal dialekter med et antal ækvivalenter.

Om strukturen i begrebsknuden til højre kan løse mine problemer, kan jeg i skrivende stund ikke gennemskue, men jeg arbejder videre med problemstillingen i mit ph.d.-projekt.

9. Sammenfatning

Således kan jeg sammenfatte at med et materiale som *Ømålsordbogens* hvor ord som *mælkebøtter*, *flødeboller* og *smørkringler* betyder noget helt andet end man skulle tro – simpelthen fordi det meste af den verden materialet handler om, er mere eller mindre ukendt i vore dage – så vil en systematisk/onomasiologisk gennemgang af materialet være af væsentlig betydning når man sidder med den alfabetiske ordbogsredaktion.

Et begrebsnetværk i lighed med DanNet kan sandsynligvis være en hjælp i arbejdet med at få struktureret data yderligere og få overblik over dem. Men det kræver nogle justeringer. Det er – om noget – den geografiske variation vi arbejder med på *Ømålsordbogen* – så den *skal* indarbejdes i de værktøjer vi skal arbejde med.

KILDER OG LITTERATUR

Andersen, Lena Wienecke, under udarbejdelse: Fra virkelighedens verden til ordbogens.

Et studie i datastrukturering med henblik på betydningsdefinitioner i ordbøger.

Igangværende ph.d.-projekt på *Ømålsordbogen*, Afdeling for Dialektforskning, Nordisk Forskningsinstitut, Københavns Universitet.

Bergenholtz, H., I. Cantell, R.V. Fjeld, D. Gundersen, J.H. Jónsson og B. Svensén, 1997: *Nordisk Leksikografisk Ordbok*. Oslo.

DanNet: <http://wordnet.dk> (maj 2011).

Geeraerts, Dirk, 2010: *Theories of Lexical Semantics*. Oxford.

Møller, Kristen, 1938: Mælken og dens Produkter. I: *Danske Folkemaal* 12. S. 1–16.

Pedersen, B.S., A. Braasch, S. Nimb, J. Asmussen, N. Sørensen, H. Lorentzen og L. Trap-Jensen, 2010: *Lingvistiske specifikationer for DanNet Version 1.0.2*. <http://wordnet.dk> (maj 2011).

Skjelborg, Åge, 1967 (revideret afskrift): *Topografisk ordning af Danmarks egne med alfabetisk nøgle*. København.

¹ At vi så i *Ømålsordbogen* vælger at generalisere til for eksempel S(alm) om ord der er almindeligt forekommende på Sjælland, L-Fa(spor opt) om ord der er sporadisk optegnet på Lolland og Falster og så videre, er en anden sag.

- Svensén, Bo, 2004: Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik. Stockholm.
- Ømålsordbogen. En præsentation, udg. af Afdeling for Dialektforskning, Københavns Universitet, 2004. København.
- Ømålsordbogen. En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omliggende øer, red. af Afdeling for Dialektforskning, Københavns Universitet, udg. af Universitets-Jubilæets Danske Samfund, 1992ff, bind 1ff. København.
- Ømålsordbogens samlinger. Afdeling for Dialektforskning, Københavns Universitet. København.
- Ømålsordbogens tillægsbind, red. af Afdeling for Dialektforskning, Københavns Universitet, udg. af Universitets-Jubilæets Danske Samfund, 1992. København.

Lena Wienecke Andersen

Ømålsordbogen, Afdeling for Dialektforskning, Københavns Universitet.
wienecke@hum.ku.dk