

Sproglige problemer i informationsvidenskabeligt og biblioteksfagligt arbejde

Risager, Karen; Lau, Birgitte

Published in:
Roskilde Universitetscenter. Lingvistgruppen. Rolig-Papir

Publication date:
1979

Document Version
Tidlig version også kaldet pre-print

Citation for published version (APA):
Risager, K., & Lau, B. (1979). Sproglige problemer i informationsvidenskabeligt og biblioteksfagligt arbejde. *Roskilde Universitetscenter. Lingvistgruppen. Rolig-Papir*, 16, 45-58.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy

If you believe that this document breaches copyright please contact rucforsk@ruc.dk providing details, and we will remove access to the work immediately and investigate your claim.

Birgitte Lau og Karen Risager

Sproglige problemer
i informationsvidenskabeligt
og biblioteksfagligt arbejde

1. Indledning.

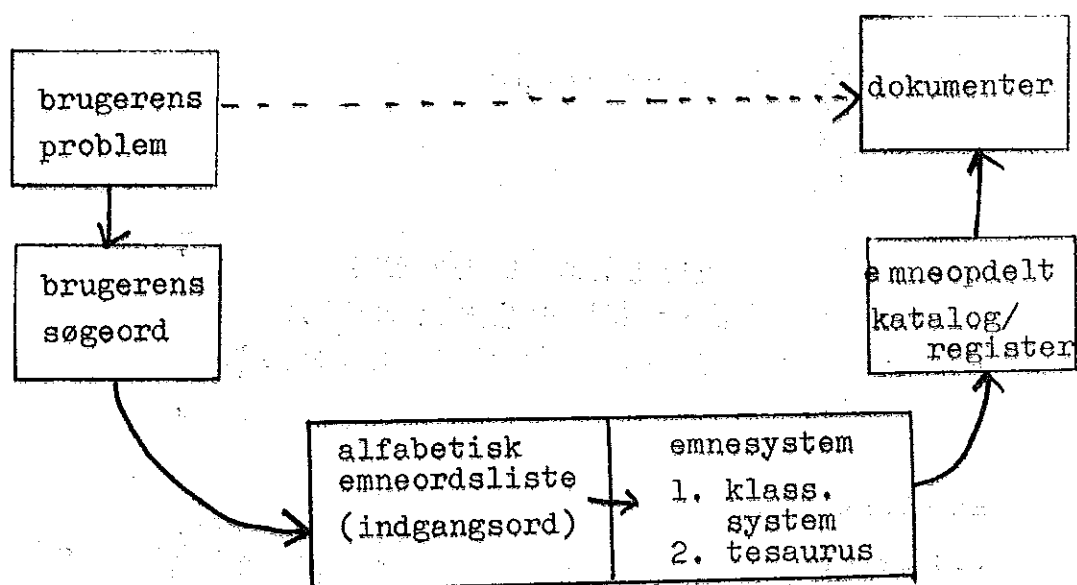
Det følgende er et forsøg på at give et overblik over nogle områder indenfor biblioteks- og informationsarbejdet hvor det efter vores opfattelse ville være oplagt at inddrage sprogvidenskabelig viden, eller hvor det ville være nyttigt om sprogfolk deltog i løsningen af problemer som i øjeblikket løses på utilfredsstillende måde. Vi vil især pege på to områder:

A. Vi mener sprogfolk kunne bidrage til at opbygge den semantiske side af hvad vi kalder emnesystemet i det følgende, og vi anser det for at være vigtigt at det gøres på en sådan måde at informationssøgningens kognitive problemer i det omfang det er muligt drages med ind i arbejdet.

B. Vi mener at sprogfolk ville være de rette til at afdække og analysere de sprogpoltiske problemer som følger af det stadig større pres på de små sprog som videnskabelige sprog.

2. Brugerens søgeproces.

Den "omvej" brugeren må gå via bibliotekernes emnesystemer for at finde dokumenter der belyser hans/hendes problem, kan illustreres ved følgende, forenkede model:



Det der har den største interesse i denne sammenhæng er det sted hvor brugeren v.h.j.a. sine søgeord forsøger at komme ind i systemet. Det er her de fleste frustrationer over biblioteker opstår, og det er her man først og fremmest skal sætte ind. Vi har en idé om at en stor del af disse frustrationer er baseret på temmelig banale sproglige kommunikationssvigt, der for en stor del skyldes forskellige leksikaliseringer hos system og bruger. Men i øvrigt er hele den proces en bruger gennemgår ved informationssøgningen sprogligt forankret, ikke bare sådan at forstå at han/hun sideløbende taler om det der sker, f.eks. med bibliotekaren, men faktisk sådan at forstå at de ord man vælger til at søge med, altså søgeordene, også udgør selve handlingen man foretager, altså er en slags sproghandlinger. Om kommunikationen lykkes eller ej afhænger af om man har brugt de rigtige begreber.

For at lette overgangen mellem bruger og søgeord og bibliotekets emnesystem har man de fleste steder et hjælpemiddel der fungerer som en slags "switching language", i form af en alfabetisk emneordsliste, der så at sige sluser brugeren fra sine egne begreber over i emnesyste-

mets. At foregribe brugerens begreber ved opbygningen af en sådan emneordsliste er en af de vanskeligste opgaver i biblioteksarbejdet; det er desværre et hyppigt forekommende tilfælde, at brugerens begreber ikke findes i listen, selv hvor biblioteket har forsynet kategorierne i emnesystemet med flere alfabetiske indgange. Disse problemer kan illustreres ved følgende eksempler:

Brugerens søgeord	Alf. emneord	Kategorier i emnesystem
ortografi	→ ortografi	→ ortografi
sprognormer	→ sprognormer	→ sprogrigtighed
sprogsocialisering	%	
%	sprogudvikling	→ sprogudvikling
%	sprogpedagogik	→ sprogpedagogik
%	børnesprog	→ sprogudvikling
%	børnesprog	→ alders- og kønssprog
kode	----→ kode	→ cybernetik
%	klassesprog	→ sociolingvistik
%	sociolingvistik	→ sociolingvistik
læsbarhed	----→ læsbarhed	→ anvendt lingvistik
%	sprogforståelse	→ sprog- og tankeprocesser

----→ = uønsket henvisning

Det grundlæggende problem er at biblioteksfolk arbejder fra højre til venstre i modellen, mens brugere arbejder fra venstre mod højre, og at det, især hvad angår nye fagtermer, er yderst vanskeligt at sikre sig at emneordslisten er ajour med den nyeste forskning.

3. Forskellige typer af emneregistreringssystemer.

Hvis vi går lidt tættere på det vi ovenfor har kaldt emnesystemet, kan man konstatere at det kan dække et bredt spektrum af typer af emneregistreringssystemer, hvorefter vi skal fokusere på to typiske former: Det klassiske hierarkisk opbyggede klassifikationssystem og det verbale emneindekseringssystem. Det hierarkiske klassifikationssystem er baseret på en opdeling i entydigt afgrænsede klasser af dokumenter. Hver klasse kan så inddrages i underklasser, o.s.v., således at betydningsindholdet i en given kategori eller klasse kan aflæses af placeringen i forhold til de øvrige niveauer i hierarkiet. Et typisk eksempel er decimalsystemet, der f.eks. er kendt fra folkebibliotekerne, hvor én decimal tilføjes, hver gang man bevæger sig ned på et nyt niveau i systemet. Iøvrigt er det også karakteristisk, at hierarkiske klassifikationssystemer anvender ikke-sproglige notationer til at markere kategorierne.

Den måde man normalt anvender klassifikationssystemet på når man står med det enkelte dokument, er at placere det sammen med beslægtede dokumenter i en slags ækvivalensgruppe, hvor alle dokumenter har samme kategorimærke. Emneindekseringssystemerne derimod bruges oftest til at beskrive det enkelte dokument v.h.j.a. en række termer (ofte kaldet deskriptorer), som tilsammen udtrykker dokumentets indhold, men som hver for sig kan udgøre en indfaldsvinkel til dokumentet i en søgesituation, alt efter hvilke termer brugeren har som udgangspunkt.

Klassifikationssystemer og emneindekseringssystemer har i de senere år i stigende grad været opfattet som konkurrerende typer af emnebeskrivelser for dokumenter, selv om det ofte er blevet fremhævet, at der her ikke er tale om et enten-eller, men et både-og. På mange måder kan de to typer af systemer nemlig siges at supplere hinanden - hvilket er nogenlunde det samme som at sige at hver type

af system har visse fordele som det andet ikke har: Klassifikationssystemer er f.eks. generelt bedre egnede som opstillingssystemer for bøger, mens indekseringssystemer er klassifikationssystemerne overlegne m.h.t. muligheder for informationssøgning i grænseområder eller for emner der går på tværs af etablerede forsknings- og faggrænser. Emneindekseringssystemer varierer betydeligt i deres opbygning, med betydelige konsekvenser for de søgemuligheder de indebærer. Nogle af de vigtigste variable er følgende:

1. Hvordan vælges indekseringstermer til et dokument?

Man kan her skelne mellem tre hovedtyper: a. Systemer der kun tillader at man vælger termer fra en kontrolleret liste i form af en thesaurus. b. Systemer der (på automatiseret vis) indekserer dokumentet v.h.j.a. alle de i titlen/abstractet forekommende ord, minus småord osv. som er frasorteret ved en stopliste. c. Systemer der tillader brug af et åbent vokabular i den forstand at indekseren kan vælge termer til beskrivelse af dokumentet som hverken indgår i thesaurus'en eller titel/abstract.

2. Hvilken opbygning har den enkelte gruppe af indekseringstermer, knyttet til ét dokument? (indførelsens opbygning).

Den simpleste form for struktur forekommer ved at lade termene optræde som enkeltstående, sideordnede indgange til dokumentet, og så lade sammenhængen være et produkt af brugerens selektion og kombination af termer i en søgeprofil:

A:	Dokument 1 har termerne:	a,b,c,d
fri	2	a,c,d,e,h
kombina-	3	a,d,e
tion	4	b,c,e,g

Brugeren har f.eks. en søgeprofil der går ud på at han/hun vil have dokumenter med følgende termer:

a + (b eller c)

Denne profil vil give dokumenterne 1 og 2.

B: Dokument 1 har tre termer: a,b,c
koordi- Ud fra disse termer produceres følgende
neret indførsler for dokumentet:
a b c
b c a
c a b

O: Dokument 1 har f.eks. tre termer, hvis
"syn- indbyrdes "syntaktiske" roller forsøges
taktisk" udtrykt, så misforståelser undgås.
Termerne kan f.eks. være:
sprogundervisning, indfødte, engelsk
Indførslerne vil blive konstrueret således,
at det fremgår om det er sprogundervisning
af eller ved indfødte, og om det er undervisning
i eller på engelsk.

3. Hvordan er de enkelte termer relateret til hinanden i systemets overordnede struktur?

Det kan godt lade sig gøre ikke at have nogen anden ordning af indekseringstermerne end en rå alfabetisk liste, oven i købet uden kontrol med stavemåder eller numerus på substantiver. Men det lægger en urimelig arbejdsburde på brugeren m.h.t. synonym- og ordforms-kontrol. Tesauri er meget nyttige arbejdsredskaber til aflastning for dette arbejde, på den måde at de udgør et slags skelet for indekseringssystemets terminologi. Den terminologiske kontrol de udøver (ud over den kontrol der ligger i om en given term er medtaget i tesaurus'en eller ej) ligger dels i et princip om foretrukne termer, en stramning af vokabularet om færre men mere prægnante termer (ordet deskriptor bruges ofte & netop denne forstand). Dels ligger kontrollen i en form for indbyggede minihierarkier i form af angivelse af hvad der i forhold til en given term er bredere termer BT (= broader term) og snævrere termer NT (= narrower term), plus en rodekasse af termer, der samles under betegnelsen RT (= related term), som dækker over et utal af andre tænkelige semantiske relatio-

ner mellem termer.

Som helhed kan man tale om at de forskellige typer af systemer fordeler sig på et kontinuum der går fra systemer med et minimum af vokabular- og kategoristyring til systemer med et maximum af styring. Kontinuet ser nogenlunde sådan ud:

min			max
friterm søgesystemer	thesaurus + fritekst	koordineret indeks.system	monohierarkisk klass.system

Hovedproblemerne i hver ende af kontinuet er i grove træk dem, at i de minimalt styrede systemer skal brugerne i stort omfang selv kontrollere synonymmer, får altså ingen vejledning i systemets terminologi, mens man i de maksimalt styrede systemer til gengæld ofte opererer på basis af opfattelser af de forskellige vidensområder som ikke svarer til brugerens og derfor vil være utilgængelige.

4. System vs. bruger.

Vi vil nu se på hvad der karakteriserer den situation brugeren er i, hvilke behov han/hun har, og hvordan disse kan eller ikke kan opfyldes af systemet.

Umiddelbart kan man definere en forskel mellem system og bruger, som går på de forskellige abstraktionsniveauer de befinder sig på. Systemet er altid udarbejdet i et totaliserende perspektiv, konstruktørerne har villet forudse og forstrukturere potentielle brugeres potentielle vidensbehov. Brugeren vil altid betragte sit eget problem, sit konkrete vidensbehov, under sin egen synsvinkel og i sin egen praksissammenhæng. Man kan sige at der på denne måde er indbygget en modsætning mellem system og bruger; de stiller ikke sammenfaldende krav.

Som første skridt i søgningen står brugeren over for den opgave at udtrykke sit problem i en række søgeord, som

kan bruges i den videre søgning. Opgaven består i at begrebsliggøre aspekter af problemet eller emneområdet, en proces der rummer en række delvist sproglige problemer i forbindelse med strukturering af et emne: afgrænsning af emnets genstandsråde, metode, faglige tilhørsforhold, tid og sted o.a., med alt hvad dette indebærer af hierarkiseringer, overlapninger og andre semantiske relationer. En væsentlig del af de sproglige problemer i denne forbindelse er af leksikologisk art. Det er problemer der har at gøre med forholdene mellem brugerens dagligsprog og de fagsprog han eller hun må bruge; problemer med de forskellige fagsprogs terminologier, specielt dem der udspringer af at fagsprogene er udviklet til forskellig tid, inden for forskellige metodiske og teoretiske rammer, og delvist udviklet sig ved forgrening af tidligere, og mindre omfattende, terminologiske systemer; naturligvis alt sammen som følge af videnskabsområdernes udvikling. Et overblik over disse problemer i den konkrete søgeproces stiller ret store krav af kognitiv psykologisk art, og rummer samtidig en række pædagogiske problemer, hvis vi ser på hvordan et sådant overblik kunne formidles mere systematisk via en undervisning i informationssøgning.

Som næste skridt skal brugeren anvende sine søgeord som indgang til søgesystemet. Den førnævnte modsætning mellem system og bruger vil her i mange tilfælde vise sig som et sammenstød mellem brugerens søgeord og deres indbyrdes indholdsmæssige strukturering på den ene side, og systemets indholdsmæssige struktur på den anden side. Det er ret symptomatisk at litteraturen om emnesystemerne næsten udelukkende handler om systemernes formelle struktur, bl.a. om efter hvilke regler man fylder data på og får data ud, mens den side af sagen der har at gøre med systemernes indhold og som i sidste ende er afgørende for om det brugerne får ud af systemet er fagligt relevant, hidtil har været temmelig upåagtet. Med systemernes indhold mener vi de termer man kan søge på og det henvisningssystem der gør at man opdager betydningsmæssige forbindel-

ser mellem termer, hvor der findes sådanne forbindelser. Og man er først for nylig begyndt af interessere sig for hvordan brugere sprogligt-kognitivt fungerer i en informationssøgningssituation. Det lader sig let demonstrere at klassifikationssystemer med en traditionel hierarkisk opbygning, af den type som stadig næsten er enerådende inden for det danske biblioteksvæsen, forudsætter stor forhåndsviden hos sine brugere, herunder viden om tidligere tiders opfattelser af videnskabernes indbyrdes forhold, og om det verdensbillede der ligger til grund for opdelingen i vidensområder. For at forstå selve hovedopdelingen i et system som UDK, det universelle decimal-klassifikationssystem, må man f.eks. være klar over, at man på det tidspunkt hvor systemet blev udviklet, opfattede psykologi som en underdisciplin til filosofi, for formelt optræder psykologi som en sådan underdisciplin. Bl.a. som en funktion af disse forældede træk rummer UDK mange ideologisk fordrejede undergrupperinger, f.eks. narkotikamisbrug placeret som en undergruppe under etik eller moralfilosofi.

De ideelle betingelser for en søgeproces forekommer os at være en situation hvor brugeren har et så bredt spektrum som muligt af descriptorer, således at flest muligt er i overensstemmelse med hans/hendes søgeord. Desuden bør der være et meget udbygget system af krydshenvisninger mellem descriptorer. Men jo mere systemet er konstrueret efter sådanne principper, jo tydeligere bliver konflikten med systeminterne krav om systemets egen opretholdelse, krav som bl.a. vil indebære at man fra konstruktørspunktet må foretrække veletablerede fagtermer frem for nye og uprøvede, og veletablerede vidensstruktureringer.

Under søgeprocessen er det meget vigtigt at brugerens egen bevidsthed om emnets struktur og praksissammenhæng fastholdes. Systemet må ikke fremtvinge en glemsel af søgeprocessens udgangspunkt. Hvis denne fastholdelse ikke

sker, vil socialiseringen til systemet i det lange løb blive farlig, idet man vænner sig til kun at finde den slags søgning relevant, som søgesystemet er konstrueret til. Og søgesystemerne er som nævnt ovenfor i stort omfang konstrueret ud fra implicitte verdensopfattelser med forskellige ideologiske fordrejninger.

En række særlige problemer opstår hvis søgeprocessen går via et on-line søgesystem, dvs. et system hvor brugeren selv v.h.j.a. manipulationer ved en dataterminal søger på et datamatisk indkodet lager af dokumentbeskrivelser. Specielt hvis dette system har indbygget mulighed for mere eller mindre avanceret dialog mellem bruger og system.

En række delvist beslægtede problemer findes i forbindelse med kommunikationen mellem bruger og bibliotekar. Det drejer sig om dels de magtforhold der er som følge af parternes forskelligartede viden og forskellige muligheder for adgang til viden. Dels forskellige forventninger fra brugers og bibliotekars side om hinandens funktioner. Dels problemer med at gøre konkrete vidensbehov forståelige og oversættelige til søgesystemet. I denne forbindelse er det undertiden et problem at bibliotekarer ofte er socialiseret til systemet og dets iboende abstrakte vidensorganisering, samt dets biblioteksfaglige terminologi, mens brugeren er socialiseret i andre retninger; derved kan der opstå nogle barrierer i kommunikationen som er vanskelige at overskride.

5. Sproglige konsekvenser af systemernes institutionalisering.

Det er vigtigt at fastslå og nøjere analysere søgesystemernes socialiserende potentiale gennem længere tids brug. En del af denne socialisation foregår ad sproglig vej. Vi skal i det følgende kort signalisere nogle mere sproglige konsekvenser af de omfattende, institutionaliserede søgesystemer der anvendes i forbindelse med snart sagt

alle biblioteker og dokumentationscentre.

Der er for det første et forældelsesproblem. Et søgesystem der er konstrueret med kontrolleret vokabular, og ikke med fri tekstsøgning, vil uundgåeligt blive forældet, indeholde forældede ord og forældede faglige strukturer. Dette tvinger brugeren til en slags oversættelsespraksis mellem forældede og aktuelle ord. Et sådant system er i sig selv konserverende og formidler en delvis forældet vidensstruktur. Man kan kæmpe mod denne tendens ved at udarbejde tesauri der indeholder kronologiske lag af termer, og som har krydshenvisninger mellem ældre og nyere termer. Dette er dog et temmelig stort arbejde, hvis tesaurusen virkelig skal være up to date; cf. arbejdet i sprognævn der omfatter lignende problemer. Man kunne vælge at lægge mere vægt på fri tekstsøgning; derved ville man få fat i flere aktuelle termer. På den anden side ville dette i virkeligheden rumme et endnu større forældelsesproblem fordi ordvalget i fx. titler ofte er stærkt bundet til bestemte, kronologisk afgrænsede brugssammenhænge. Man er derfor her nødt til at finde en balance.

For det andet er der et monopoliseringsproblem. Et kontrolleret vokabular betyder udvikling af en kanoniseret del af ordforrådet, hvilket kan være ensbetydende med en favorisering af bestemte videnskabsretningers ordforråd. Da søgesystemerne har en monopolstilling i formidlingen af viden, fungerer de som sprogundertrykkere og undertrykkere af bestemte slags viden. Også her kan man kæmpe mod denne tendens ved at udarbejde større, og fagligt bredere tesauri.

Man kan altså sige at med henblik på anvendelse af sprogvidenskabelig viden er arbejdet med tesauri en meget central aktivitet i forbindelse med forbedring af søgesystemerne, og forbedring af brugernes muligheder for at fremskaffe fagligt relevant materiale.

6. Centralisering og internationalisering.

De sidste ti år er der blevet udviklet meget omfattende on-line søgesystemer, som giver mulighed for at søge på databaser der indeholder flere millioner dokumentbeskrivelser. Sådanne systemer har naturligvis en enorm magtposition jo mere omfattende og centraliserede de bliver. Og jo mere bliver den sprognormerende og sprogpolitiske karakter af tesaurusarbejdet tydelig, omend denne karakter ofte legitimeres teknokratisk.

Allerede med de store internationale bibliografier har man set en monopolisering af nogle enkelte hovedsprog, først og fremmest engelsk. Der er her tale om en slags international sprogkamp, hvori stadig flere bibliografier er inddraget, også efterhånden de mindre og lokale. Udviklingen betyder bl.a. at det er den engelsksprogede videnskabsstrukturering og de engelsksprogede ideologiske strukturer i almindelighed som litteraturen er organiseret efter.

For os der har med dansk at gøre, har denne udvikling en aktuell interesse. Inden for EF-regi og UNESCO-regi er der efterhånden flere planer om at videreudbygge de allerede delvist eksisterende net af databaser, i Europa f.eks. Euronet, og at udarbejde multilingvale tesauri til hjælp for søgningen i disse net. Udover de allerede nævnte problemer med forældelse og monopolisering af bestemte dele af ordforrådet, ligger der her et oplagt sprogpolitisk problem: Skal dansk overhovedet inkorporeres i disse tesauri? Overvejelserne må tage hensyn til bl.a. hvem i Danmark der skal kunne have adgang til nettene: Fagfolk med tilstrækkeligt kendskab til de fremmedsprog der under alle omstændigheder vil blive taget med i disse tesauri, eller den brede befolkning? Da det er tvivlsomt om fremmedsprogsundervisningen i Danmark vil blive tilstrækkelig effektiv til at sætte alle i stand til at bruge fremmedsprog i søgningen, vil

en udelukkelse af dansk på nuværende tidspunkt betyde at store dele af den danske befolkning på forhånd vil være ude af stand til at bruge systemerne. Desuden vil det antagelig blive vanskeligere at søge bestemte emner, nemlig dem der er specifikt danske (opstået i Danmark, og udelukkende tematiseret på dansk). Herudover må man overveje hvilke konsekvenser det kan have på længere sigt for dansk som videnskabeligt sprog, hvis man nu bestemmer sig for at udelukke dansk som udgangspunkt for søgning. Formodentlig vil tendensen til at skrive arbejder på engelsk blive stærkere, med de konsekvenser dette så kan have for kommunikationen mellem fagfolk og den brede befolkning herhjemme.

Disse sprogpolitiske problemer vil vokse i takt med den voksende internationalisering og centralisering i Europa og i verden. Ofte er de politiske afgørelser dikteret af økonomiske og tekniske krav, og sådan bliver de som regel også begrundet. Men de sprogpolitiske problemer drejer sig også om magt og undertrykkelse, og det forekommer os at være en vigtig opgave for sprogvidenskaben at analysere sådanne problemer nøjere og udvikle en holdning til i hvert fald de sproglige konsekvenser af de politiske beslutninger.

BILAG:

Oversigt over behov for sprogvidenskabelig bearbejdelse:

1. Semantiske teorier, især omkring det leksikalske niveau (f.eks. synonymforskning, feltteori) og det syntaktisk-semantiske niveau (f.eks. kasusteori)
Til: udvikling af tesauri m.h.t. beskrivelse af semantiske relationer mellem termer.
Til: beskrivelse af relation mellem problem og søgeord.
Der er især brug for forskning i leksikologisk semantik og sociosemantik, bl.a. vedr. forholdet mellem daglig-sprog og fagsprog.
Til: opdatering af tesauri m.h.t. nye fagtermer.
Der er også brug for forskning i oversættelsesproblemer.
Til: udarbejdelse af multilinguala tesauri.
2. Sprogpsykologi, især omkring begrebsstrukturer, kognitive mønstre.
Der er brug for forskning i "relevansstruktur", og i det hele taget for mere socialisationsorienterede angrebsvinkler ved beskrivelser af begrebsstrukturer.
Til: undersøgelser af brugernes kognitive mønstre i informationssøgningsituationen.
Til: undersøgelser af den sproglige socialisationsproces i brugen af informationssystemet.
3. Samtaleanalyse/-teori, især omkring interviews mellem serviceinstitutioner og brugere.
Til: undersøgelser af kommunikationsproblemer mellem bruger og bibliotek.
Til: undersøgelser af hvordan emner defineres og afgrænses.
4. Artificial intelligence
Til: udvikling af bedre on-line dialogsystemer til edb-informationssøgning.
5. Sprognormering og sprogpolitik
Til: tesaurusarbejde, incl. på internationalt plan.
Til: modarbejdelse af uønskede politiske konsekvenser af sprognormering.

ROLIG papir

Karen Risager og
Niels Haastrup
(red)

Om
anvendelse
af
sprogvidenskabelig
viden
i
erhvervsfunktioner
uden for
skolesystemet

Rapport fra den
4. studiekonference
i nydansk grammatik
og sprogbeskrivelse

Roskilde Højskole
6.-8. april 1979

16 | **79**

ROSKILDE UNIVERSITETSBIBLIOTEK

20. 09. 79
Periodica afd.

Roskilde UniversitetsCenter
Lingvistgruppen

Rol

ROSKILDE UNIVERSITETSBIBLIOTEK



TIDSSKRIFTER

370006728407

ex 5