

## Rapport om arbejdet med metataksnomien i pilotfasen

Rapport om arbejdet med metataksnomien i pilotfasen .....	1
Resume .....	1
Status .....	1
Ortografi .....	1
Morfologi .....	2
Semantik .....	2
Tidsforbrug .....	2
Erfaringer og anbefalinger til næste fase .....	2
Hvor meget mangler der? .....	4
Oversigt over bilag .....	4

### Resume

Arbejdsgruppen har nået næsten at færdiggøre området ortografi. Det indeholder 39 begreber/kategorier. Der er endnu ikke fuld enighed om strukturen i den hierarkiske taksonomi, men det forventes at vi bliver enige når vi ses i Borgå.

Pilotfasens øvrige områder, begreber under semantik og morfologi, har alle fået en engelsk definition, men i Sverige og Norge har der ikke været tilstrækkeligt tid til at tilføje nationale termer og til at kommentere de engelske definitioner. Starten blev forsinket i både Sverige (blandt andet pga. sygdom) og Norge (fordi der manglede personer til opgaven).

Arbejdsgruppen har undervejs diskuteret hvordan definitionerne skal formuleres. Når gruppen mødes i Borgå, vil den kunne nå til en fælles forståelse af principperne. Derefter vil arbejdet sandsynligvis kunne gå hurtigere.

### Status

Pilotprojektets konsensusproces har i praksis foregået i perioden 23. juni til 19. august.

Den danske deltager og koordinator gik for alvor i gang den 15. maj.

Den svenske deltager gik for alvor i gang omkring den 23. juni.

De to norske deltagere gik for alvor i gang omkring den 3. juli.

I pilotprojektet har det været hensigten at følgende emner skulle udfyldes med engelske definitioner, danske, svenske, norske termer og synonymer. Desuden skulle arbejdsgruppen blive enige om strukturen i taksonomiens hierarkiske opbygning, de engelske termer, synonymer og definitioner.

I løbet af ovennævnte periode har arbejdsgruppen nået følgende.

### Ortografi

#### Status

Der er 46 artikler under dette emne, 7 af disse ligger uden for pilotprojektet, og der er altså 39 artikler.

For dansk og svensk er der skrevet en national term (og synonymer) for alle disse. Desuden har alle deltagere kigget på strukturen og den engelske definition.

For norsk bokmål er der tilføjet termer i 19 artikler og ikke for 26 artikler.

For nynorsk er der tilføjet termer i 18 artikler og ikke for 27 artikler.

### **Forklaring på at der mangler noget**

Hovedårsagen til at gruppen ikke har nået det hele, er at arbejdet er kommet for sent i gang. Der er ikke i alle artikler givet en acceptkode fra alle. Det er altså ikke muligt med sikkerhed at se om de pågældende har kigget på den engelske definition. Dette skyldes dels den generelle mangel på tid, og dels at det er svært at opnå konsensus om termerne før strukturen er etableret. Der har været diskussion om strukturen undervejs, men der er ikke blevet opnået fuld enighed.

Koordinatoren (IEM fra Dansk Sprognævn) vil forberede et oplæg til mødet i Borgå således at arbejdsgruppen kan komme frem til en løsning på strukturproblemet.

Når der er opnået klarhed over strukturen, vil det ikke kræve meget tid at opfylde målsætningen for emnet ortografi. Det vil kræve mest af nordmændene som mangler mest.

### **Morfologi**

Der er 52 artikler som alle ligger inden for pilotfasen.

For engelsk er der tilføjet synonymer og forslag til definition til alle.

For dansk er der tilføjet termer, synonymer og definitioner for alle.

For svensk er der tilføjet termer for ca. 7 artikler.

For norsk er der ikke tilføjet termer.

### **Semantik**

Der er 12 artikler som alle ligger inden for pilotfasen.

For engelsk er der tilføjet synonymer og forslag til definition til alle.

For dansk er der tilføjet termer, synonymer og definitioner for alle.

For svensk er der tilføjet termer for 3 artikler.

For norsk er der ikke tilføjet termer.

Der er ikke andre forklaringer på forsinkelsen inden for morfologi og semantik end at arbejdsgruppen har haft for lidt tid.

### **Tidsforbrug**

Koordinator, IEM: 137 timer (april- 31. august)

UKD

DI

ØB

### **Erfaringer og anbefalinger til næste fase**

#### *Erfaring om arbejdsformen*

Arbejdsformen har fungeret ok. Alle har brugt mailen på en god måde. Når vi går videre, kan vi blive endnu bedre til at bruge kommentarfelterne. Det er en ulempe i i-Term at man ikke automatisk

kan søge de artikler frem som indeholder en ny kommentar. Men dette er primært koordinatorens problem.

Alle har skullet lære at bruge i-Term først. Til dette har koordinatoren udarbejdet en *vejledning i brugen af i-Term* (se bilag 1) – og den er blevet revideret undervejs i samråd med arbejdsgruppen.

Sommerferien har forstyrret flowet i processen fordi gruppen har haft forskudt ferie. Men det kan ikke undgås om sommeren.

#### *Anbefaling om arbejdsformen*

I projektets videre færd kan man forsøge at aftale nogle bestemte uger hvor alle arbejder på et bestemt område i taksonomien. Det er nemlig vanskeligt at komme med en dybdegående analyse og gode argumenter for eller imod noget som man ikke selv aktuelt sidder og arbejder med.

#### *Erfaring om projektets formål og vision*

For deltagerne i arbejdsgruppen har der fra tid til anden været uklarhed om projektets overordnede mål, og om metataksnomiens rolle i en fremtidig søgeportal.

#### *Anbefaling vedrørende projektets formål og vision*

Det er vigtigt at sørge for kontinuitet i bemanningen – at de samme personer så vidt muligt fortsætter.

Det er vigtigt at arbejdsgruppen mødes til netværksmødet i Borgå og opbygger en fælles forståelse af hvordan metataksnomien tænkes ind i en kommende søgeportal. I søgeportalen skitseres følgende muligheder for søgning: på kategori alene, flere kategorier, kombinerede kategorier og en kombination af kategori- og fritekstsøgning.

Det vil være en hjælp fremover at blive enige om eksempler på gode definitioner i metataksnomien – og evt. problematiske. Eksempler på gode definitioner findes i bilag 2.

#### *Lingvistisk ontologi eller taksonomi til klassifikation af spørgsmål og svar*

Visse af de diskussioner arbejdsgruppen har haft skyldes uklarhed om slutproduktet: er det en lingvistisk ontologi som i den engelske referencedel er sproguafhængig, eller er det en taksonomi som har det specifikke formål at klassificere sproglige spørgsmål og svar?

Det ligger i selve projektets kerne at det egentlig primært er en taksonomi som drejer sig om spørgsmål og svar. Imidlertid er det interessant at der som et biprodukt kommer en slags lingvistisk ontologi ud af det. Problemet er bare at det i praksis er svært at skrive og enes om en engelsk definition af et begreb som både er en lingvistisk etableret term og en kategori som klassificerer sprogspecifikke spørgsmål og svar om emnet.

I bilag 3 findes en oversigt over denne og andre problemstillinger som arbejdsgruppen er stødt på. Ved en del af dem er der opnået enighed. Det er hensigten i Borgå at finde principielle løsninger på de resterende som vil gøre arbejdet lettere fremover.

#### *Målgruppe for søgeportalen*

Ovenstående hænger tæt sammen med portalens målgrupper. I første omgang tænkes der på ekspertbrugere som målgruppe. Der argumenteres i ”Resume af kravspecifikation for en nordisk svarbaseportal” for dette valg.

### ***Hvor meget mangler der?***

Grundlaget for metataksnomien er den danske DSN-emnetaksonomi. Den indeholder i alt 325 kategorier. De 88 af disse indgår i pilotprojektet, men heraf mangler ca. 60 begreber stadig termer på svensk og norsk og konsensus om de engelske definitioner.

Der resterer 237 kategorier fra den danske taksonomi, men dette antal kan øges i løbet af arbejdsprocessen. Det sker typisk fordi et land ønsker at en kategori skal have flere underbegreber. Det er også muligt at der er emner som kun er relevante i andre lande og ikke er med i den danske taksonomi.

### ***Oversigt over bilag***

Bilag 1: Vejledning i brugen af i-Term\_v2

Bilag 2: Gode eksempler på definitioner

Bilag 3: Detaljerede erfaringer i arbejdsgruppen