

NordTerm 2019

11.-14. juni, København



Proceedings og landerapporter

Redigeret af Lina Henriksen & Sussi Olsen

Center for Sprogteknologi, NorS, Københavns Universitet

Oktober 2019

ISSN 1100-9659

Finansieret af



KØBENHAVNS UNIVERSITET



Indholdsfortegnelse

Forord

Tommi Bernitz: Anvendelse af maskinoversættelse i Europa-Parlamentet: behov og muligheder for integration af terminologi

Lotte Weilgaard Christensen, Bodil Nistrup Madsen: Terminologisk ontologi for 'indsatsledelse' på dansk

Kristine Eide: Språkressurser i språkteknologien – termenens uerstattelighet og hvordan overtale offentlige etater til å ta ansvar for språkteknologi

Brian Jacobsen: A Practical Example of the Use of Terms in Automated Classification

Stine Jensen: Nordic European Terminology Collaboration and Networking

Lars Bagøien Johnsen: Digitaliserte tekster for terminologisk analyse

Susanne Lervad, Christian Gaubert: Textile Terminology in the Thread project – a multilingual approach

Henrik Nilsson: Svårt med facilitering? – om terminologiska arbetsgrupper, konsensus och terminologens roll

Henrik Nilsson: Lagringsvaror men inga lagervaror – om europeiska nationella termbankar och deras programvaror

Niina Nissilä & Anita Nuopponen: Hur mycket språkteknologi behövs det i terminologiutbildningen?

Sussi Olsen, Lina Henriksen: eTranslation TermBank, terminologi på tværs af Danmark - og EU

Katri Seppälä, Mari Suhonen: Digitalisering av offentliga tjänster – begreppsarbete som en del av utveckling och digitaliseringen av statsunderstödsverksamheten

Gerd Sjögren, Mats Granström, Henrik Andersen: When Automation Requires a Semantic Interface: Some Examples

Gunnhildur Stefánsdóttir: From Glossary to Term Base. A Tale of Revision – A Question of Quality and Need

Hanne Erdman Thomsen: Relationer i Skatteforvaltningens Informationsmodeller

Ágústa Þorbergsdóttir: Nye perspektiver og udfordringer for terminologi i en islandsk kontekst

Marit Torseth, Jim Yang, Ole Våge: Høsting av digitaliseringens frukter: Hvordan begrepskataloger kan bidra til språkteknologi

Landerapporter

Danmark

Finland

Island

Norge

Sverige

Forord

I juni 2019 skulle Nordterm-konferencen afholdes hos os på Center for Sprogteknologi ved Institut for Nordiske Studier og Sprogvidenskab (NorS), på Københavns Universitet.

Vi besluttede, at vi med temaet i år ville forsøge at tegne et billede af terminologiens fremtid i forbindelse med sprogteknologiske værktøjer og kunstig intelligens. I Danmark – og mange andre steder i øvrigt – har der de seneste år været debat og politisk interesse for emner, som på forskellig vis vedrører kunstig intelligens. Således har den danske regering i 2018 udgivet en rapport, *Digital service i verdensklasse*, som forudsætter, at Danmark bliver førende i anvendelsen af kunstig intelligens - og i 2019 blev rapporten *Dansk sprogteknologi i verdensklasse* udgivet af Det Sprogteknologiske Udvalg. Dette udvalg blev nedsat af Kulturministeriet, og udvalgets kommissorium inkluderede en undersøgelse af sprogets og sprogteknologiens rolle i kunstig intelligens samt behovet for en central termbank i Danmark.

På denne baggrund blev temaet for Nordterm-konferencen i 2019: *Terminologiens rolle i sprogteknologiske applikationer som fx virtuelle assistenter og maskinoversættelse*.

Konferencens indlæg gav på mange forskellige måder hver deres bidrag til dette tema, som det fremgår af artikler og abstracts i disse proceedings, men der er især 2 talere og præsentationer, som bør nævnes her. Det drejer sig om Sabine Kirchmeiers indlæg: *Digital service i verdensklasse. Nye perspektiver for kunstig intelligens, sprogteknologi og terminologi i den offentlige sektor i Danmark* og Andrejs Vasiljevs indlæg: *Terminology in the age of Artificial Intelligence – decline or revival?*

Sabine Kirchmeier, en af hovedtalerne, rapporterede som formand for Det Sprogteknologiske Udvalg om nogle af rapportens konklusioner. Udvalget anbefaler bl.a. etablering af en dansk sprogbank, som skal inkludere et talesprogs-korpus, opmærkede tekstkorpusser, en orddatabase og en termbank; alt sammen optimeret til sprogteknologiske formål. Det anbefales endvidere, at der i langt højere grad satses på egentlige sprogteknologiske værktøjer, men det er nødvendigt at starte med sprogbanken, fordi sproget, ordene og terminologien udgør kernen i disse værktøjer.

Det blev også understreget i Sabines præsentation, at uden en sådan sprogbank og uden sprogteknologi bliver der ingen kunstig intelligens. En robot eller virtuel assistent skal indeholde mange fakta, kunne anvende logik og ræsonnere. Den skal kunne forstå ord og sætninger, kunne fortolke – og den skal kunne udføre en ordre. Det er alt sammen noget, der forudsætter sprogteknologi, terminologi og andre sprogresurser.

Andrejs Vasiljevs, en anden hovedtaler ved konferencen, er direktør for den lettiske virksomhed Tilde, som bl.a. arbejder med terminologi, virtuelle assistenter og udvikler specialiserede maskinoversættelsesprogrammer. Tilde har gennem årene deltaget i mange EU-projekter om terminologi og oversættelse, og Andrejs pointerede, at vi i fremtiden vil have endnu større behov for store mængder terminologi, fordi terminologi er nødvendig i mange sprogteknologiske værktøjer; og kunstig intelligens kræver både terminologi og sprogteknologi. Og for Andrejs er der ingen tvivl om, at kunstig intelligens er et felt i rivende udvikling, og at vi endnu kun har set begyndelsen.

Der findes ingen artikler for Sabines og Andrejs' indlæg, men til gengæld findes der abstracts/artikler for alle de andre indlæg på konferencen. Vi ønsker jer god fornøjelse med læsningen.

Anvendelse af maskinoversættelse i Europa-Parlamentet: behov og muligheder for integration af terminologi

Tommi Bernitz

Europa-Parlamentet, GD TRAD, Den Danske Oversættelsesenhed
Luxembourg
tommi.bernitz@europarl.europa.eu

Maskinoversættelse i Europa-Parlamentet - eTranslation

Maskinoversættelse er almindeligt udbredt i dag blandt oversættere og vil sikkert vinde mere indpas i takt med, at maskinoversættelsesprogrammerne bliver bedre og mere tilgængelige for oversættere.

Europa-Parlamentet anvender ligesom andre EU-institutioner Europa-Kommissionens maskinoversættelsesprogram eTranslation til oversættelse af forskellige typer af dokumenter. Det dækker alle 24 EU-sprog med i alt 552 kombinationer. Alle oversættelser til og fra engelsk er direkte. De andre sprogpar bruger et system, hvor teksten først oversættes til engelsk og derfra til det endelige målsprog. eTranslation bruger Euramis-oversættelseshukommelser¹ som korpus. Siden 2018 har eTranslation anvendt neural maskinoversættelse.

Behov og muligheder for integration af terminologi

Maskinoversættelse har bestemt ikke mindsket behovet for terminologi i oversættelsen. Selv om maskinoversættelse, herunder neural maskinoversættelse, generelt letter oversætternes arbejde, er der stadig behov for terminologi for at sikre præcisionen og korrektheden af den terminologi, der anvendes i oversættelserne, for at kunne levere kvalitetsfyldte oversættelser.

Det er derfor meget vigtigt, at der anvendes den samme terminologi i hele oversættelsen, og maskinoversættelsesprogrammerne er ikke altid i stand til at sikre dette. Der er derfor stadig brug for, at oversættere eller terminologer sikrer, at den korrekte terminologi anvendes.

En mulighed for direkte at integrere terminologi i maskinoversættelsen vil være at fodre og træne maskinoversættelsesprogrammer med terminologi fra IATE2-termbasen² for at få bedre terminologi. En anden mulighed vil være at bruge CAT-værktøjer sammen med termbaser, der oprettes ved hjælp af term-ekstraktion til at sikre mere konsekvent og korrekt terminologi. I Europa-Kommissionen eksperimenteres der med at finde nye måder til at integrere terminologi i maskinoversættelsen.

¹ Med over 1 mia. sætninger på de 24 officielle EU-sprog, der er produceret af EU-institutionernes oversættere.

² EU-institutionernes termbase IATE - Interaktiv Terminologi for Europa.

Terminologisk ontologi for *indsatsledelse* på dansk

Lotte Weilgaard Christensen, Bodil Nistrup Madsen

Syddansk Universitet, Copenhagen Business School
Universitetsparken 1, 6000 Kolding, Dalgas Have 15, 2000 Frederiksberg
lotte@sdu.dk, bnm.msc@cbs.dk
<https://www.sdu.dk/da>, <https://www.cbs.dk/>

Abstract

I forbindelse med en artikel i en antologi inden for katastrof håndtering har vi udarbejdet en terminologisk ontologi for *indsatsledelse* baseret på en alfabetisk termliste fra Beredskabsstyrelsen. Termlisten indeholder 46 danske begreber med forklaringer. I artiklen beskriver vi principper for terminologisk arbejde og fokuserer på, hvordan man ved at strukturere information ved hjælp af den terminologiske metode kan opnå en fælles forståelse blandt aktørerne fra forskellige beredskaber og hos borgerne. Formålet med dette indlæg er at redegøre yderligere for selve arbejdsprocessen ved oprettelse af terminologiske ontologier. Med den brugerinddragende metode kortsortering er det muligt at skabe sig et første overblik over relevante begreber og deres indbyrdes relationer for derefter at udarbejde begrebssystemer som input til de mere formelle terminologiske ontologier. Vi vil diskutere og redegøre for udfordringer i arbejdsprocessen vedrørende valg af inddelingskriterier, ikke-leksikaliserede begreber og generelle begreber. I indlægget vil vi tillige beskrive og diskutere, hvordan forklaringerne i ordlisten kan omskrives til intensionale definitioner og formaliseres i form af træk-værdi-par. I en perspektivering vil vi kort berøre, hvordan terminologisk arbejde vil kunne (gen)anvendes i et grænseoverskridende samarbejde, som fundament for klassifikationer og tesaurusser og som input til andre sprogteknologiske ressourcer og applikationer.

Keywords: katastrof håndtering, kortsortering, terminologiske ontologier, ikke-leksikaliserede begreber, intensionale definitioner

1. Indledning

I forordet til Beredskabsstyrelsens publikation "Retningslinjer for indsatsledelse" (Beredskabsstyrelsen 2018:1) fremhæves det, at der i Danmark er en stærk tradition for tæt, tværgående samarbejde mellem de forskellige myndigheder og beredskabsaktører i forbindelse med beredskabsmæssig indsats. For at sikre det bedst mulige samarbejde, er retningslinjerne derfor også udarbejdet og tiltrådt af de aktører, der normalt er involveret i en sådan indsats.

I forbindelse med udarbejdelse af vores artikel til en antologi inden for katastrof håndtering (med titlen "Technological Mediation for Disaster Risk Management") var formålet at udarbejde en terminologisk ontologi inden for indsatsledelse på dansk (Christensen & Madsen 2019). Vi tog afsæt i Beredskabsstyrelsens publikation, nærmere bestemt i en alfabetisk termliste omfattende 46 centrale danske begreber, med forklaringer til begreberne (Beredskabsstyrelsen 2018:64-70). Målgruppen for publikationen er de aktører, der indgår i den samlede beredskabsmæssige indsats, herunder især politiet, redningsberedskabet og sundhedsberedskabet (Beredskabsstyrelsen 2018:7), men den vil også kunne anvendes af borgerne i almindelighed.

I indledningen fremhæves og forklares de tre centrale begreber *indsatsområde*, *skadested* og *fareområde*. Derudover nævnes de politimæssige begreber *gerningssted* og *operationsområde*. Der er her til dels tale om begreber, som også er nævnt i Aasgaards indlæg på Nordterm 2015, som er publiceret i artiklen "SPRÅKET I NØDETATER – forklare, ikke forvirre" (Aasgaard 2017). I det hele taget indeholder termlisten primært fagudtryk, som befinder sig i en gråzone mellem alment sprog og fagsprog, da udtrykkene er alment kendte af og forståelige for ikke-fagfolk, uden at disse dog kender den helt specifikke faglige betydning.

At der er behov for en terminologisk tilgang inden for området understreges blandt andet i artiklen "Crisis and Disaster Management Terminology" (2017), hvor

forfatteren Stoyan Stoyanov (som er ansat som chief expert ved CMDR COE¹, i Bulgarien) påpeger, at der blandt de forskellige institutioner inden for katastrof håndtering ikke er konsensus om definition af områdets hovedbegreber, og at disse ikke i et enkelt dokument er defineret "in clear and logically connected manner". Selv i UNISDR's² velrenommerede arbejde har han i et udkast til termliste (UNISDR 2016) konstateret inkonsistens og overlap, hvilket han igen tilskriver, at der ikke er nogen der har forsøgt at relatere begreberne til hinanden.

Formålet med nærværende indlæg er

- at redegøre for arbejdsprocessen ved oprettelse af terminologiske ontologier, herunder anvendelse af den brugerinddragende metode kortsortering
- at diskutere og redegøre for de udfordringer, vi er stødt på under udarbejdelse af den terminologiske ontologi for *indsatsledelse*, herunder
 - valg og placering af inddelingskriterier
 - inddelingskriterier vs. leksikalisering af ikke-leksikaliserede begreber
 - tilføjelse af generelle begreber i termlisten med henblik på at skabe en komplet ontologi
- at redegøre for omskrivning af termlistens forklaringer til intensionale definitioner og træk-værdi-par
- at relatere resultaterne af vores arbejde til anbefalingerne fra Sprogteknologiudvalget i Danmark.

¹ Crisis Management and Disaster Response Centre of Excellence

<https://www.linkedin.com/in/stoyan-stoyanov-b13170125/>

² UN Office for Disaster Risk Reduction

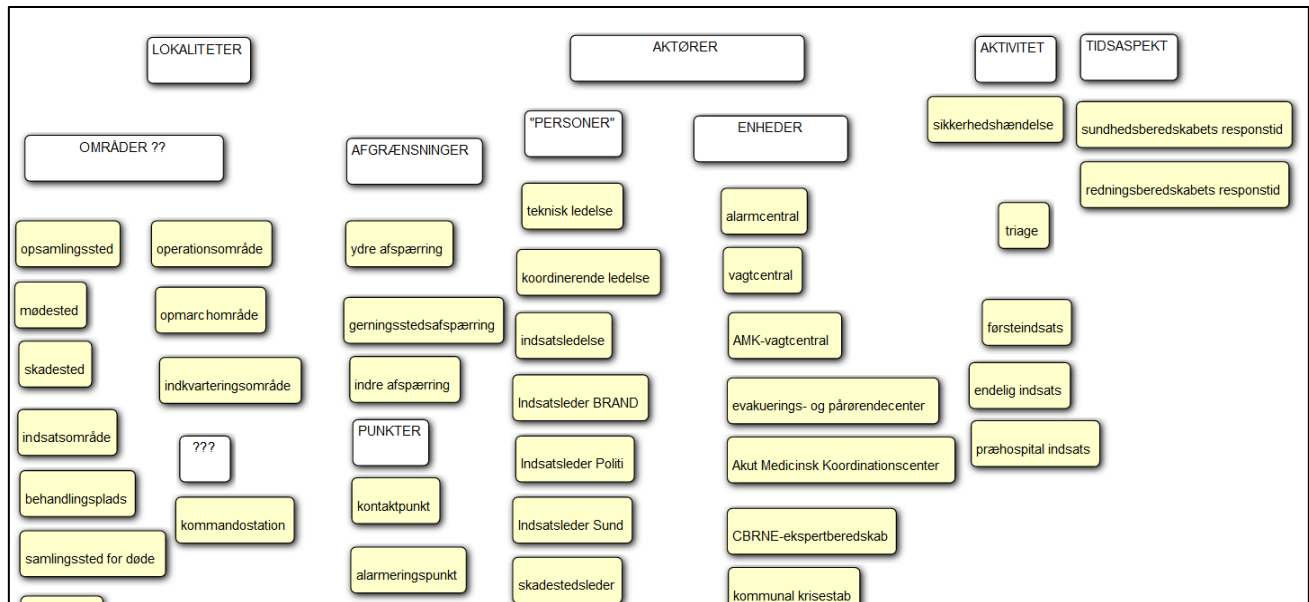
<https://www.unisdr.org/we/inform/terminology>

2. Arbejdsproces

2.1 Kortsortering

Som forstadie til at udarbejde en terminologisk ontologi over *indsatsledelse* foretog vi hver især en *åben kortsortering* ved hjælp af terminologi- og videnshåndteringssystemet i-Term (n.d.), jævnfør udsnittet i Figur 1. De gule bokse her indeholder begreber, og de hvide er kategorier. Formålet var at opnå et første overblik over begreberne og deres indbyrdes relationer.

Kortsortering er en brugerinddragende metode, der ofte anvendes af informationsarkitekter til at få indsigt i, hvordan personer forstår og sorterer information (Spencer 2009:5; Rosenfeld et al. 2015:344). Testdeltagere får udleveret et antal kort med påskrevne emner og får til opgave at inddele de udleverede kort i logiske grupper. Der findes overordnet to typer kortsortering: åben og lukket kortsortering.



Figur 1: Kortsortering (udsnit)

I en *åben kortsortering* får deltagerne udleveret kort, hvor emnerne er skrevet på kortene, men kategorierne er ikke foruddefinerede (Spencer 2009:4). Deltagerne grupperer de udleverede kort i overensstemmelse med indholdet og navngiver selv kategorierne (Rosenfeld, et al. 2015:166). I de *lukkede kortsorteringer* får deltagerne udleveret et sæt kort med emner og et sæt kort med foruddefinerede kategorier. Her bliver deltagerne bedt om at placere de enkelte kort med de påskrevne emner under de foruddefinerede kategorier (Rosenfeld et al. 2015:166; Spencer 2009:52).

Åbne kortsorteringer er egnede til indledende undersøgelser, mens de lukkede kortsorteringer er bedre egnede til validering (Rosenfeld et al. 2015:344).

Vores kortsortering af begreberne fra den alfabetiske liste mindede meget om hinanden og dannede dermed afsæt for det videre arbejde. I vores kortsortering blev kategorierne til inddelingskriterier, for eksempel AFGRÆNSNING, og til overbegreber, for eksempel *aktør*.

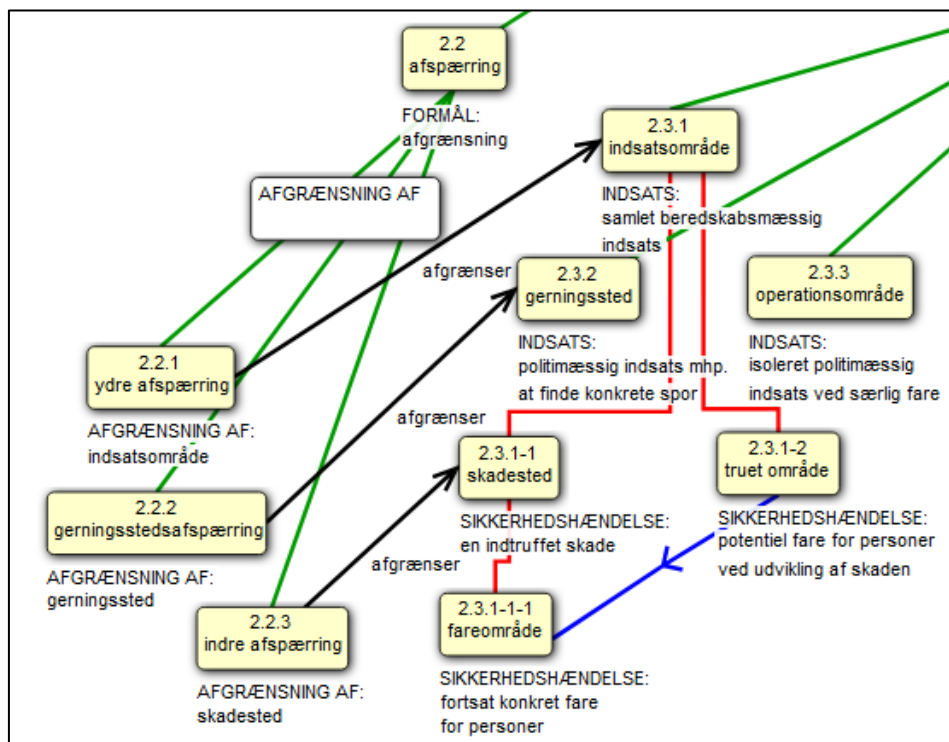
2.2 Udarbejdelse af den terminologiske ontologi

Figur 2, som ligeledes er oprettet i i-Term (n.d.), viser nogle af de begreber, som var centrale i ontologiarbejdet. Vi startede med at gå dybere ned i grupperne OMRÅDER og AFGRÆNSNINGER fra kortsorteringen.

Et af de centrale begreber er *indsatsområde*. Indsatsområdet omfatter *skadested* (med delområdet *fareområde*) og *truet område*, jævnfør de røde streger (del-helhedsrelationer). Indsatsområdet er det samlede område, hvor den beredskabsmæssige indsats foregår. Skadestedet er det område, hvor der er indtruffet en skade, og hvor der er behov for indsættelse af blandt andet redningsberedskabet. Et fareområde er en del af skadestedet, hvor der fortsat er en konkret fare for personer. Et truet område kan udvikle sig til et fareområde, jævnfør den blå pil (temporal relation).

Ud over disse områder kan der i politimæssig forstand være tale om et *gerningssted*. Gerningsstedet vil typisk have en større udstrækning end skadestedet og omfatter hele det område, hvor der er mulighed for at finde konkrete spor. Endelig optræder der et *operationsområde*, hvor der foregår en isoleret politimæssig indsats ved særlig fare.

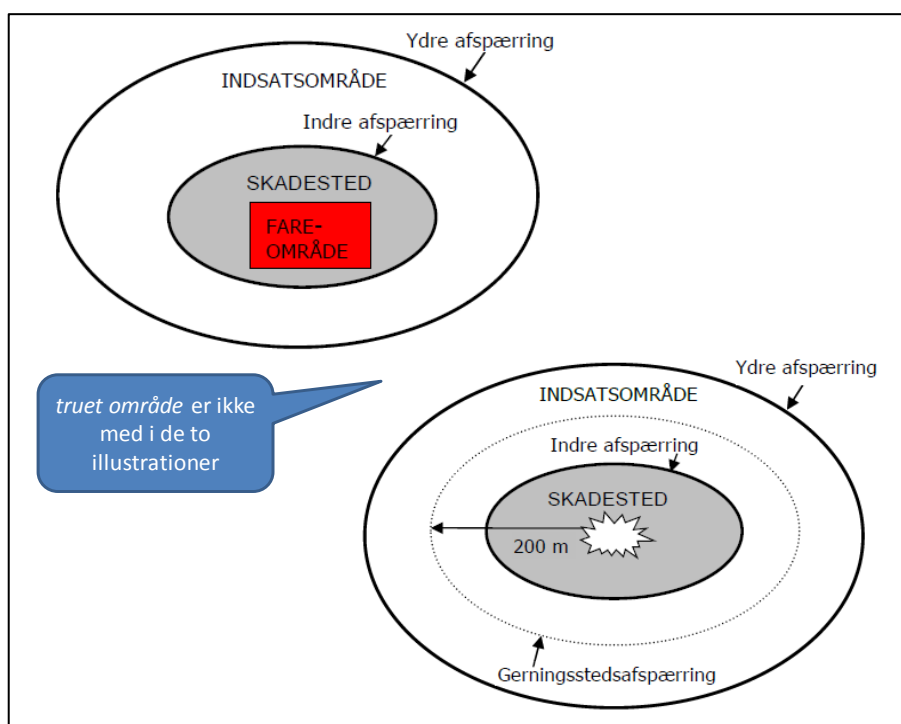
Der er tre afspærringer (*ydre afspærring*, *indre afspærring* og *gerningsstedsafspærring*), som afgrænser de enkelte områder. De indgår i typerelation med *afspærring* (de grønne linjer) og er relateret til de tre områder (*indsatsområde*, *gerningssted* og *skadested*) med associative relationer (de sorte streger med pile og relationsbetegnelser).



Figur 2: De centrale begreber i den terminologiske ontologi

Figur 3 viser to forskellige illustrationer fra Beredskabsstyrelsen (2018), hvor nogle af de centrale begreber indgår. Den øverste illustration i Figur 3 indeholder begrebet *fareområde*, mens *fareområde* ikke er eksplicit anført i den nederste illustration. Begrebet *truet område* er ikke med i de to illustrationer, da det vil være svært at illustrere: *truet område* kan befinde sig inden for *indsatsområde* eller inden for *skadested*.

I den nederste illustration er der en *gerningsstedsafspærring*, men denne mangler i den øverste illustration. Det er ikke hensigtsmæssigt, at Beredskabsstyrelsen selv har forskellige illustrationer. Den terminologiske ontologi viser tydeligt disse afspærringer og områder og deres indbyrdes relationer.



Figur 3: To forskellige illustrationer fra Beredskabsstyrelsen (2018)

3. Udfordringer

3.1 Valg og placering af inddelingskriterier

Umiddelbart kunne de områder, som vi først placerede under *område*, inddeles med samme inddelingskriterium FORMÅL. Men vi havde et ønske om at dele begreberne op i flere grupper, især så vi fik adskilt *indsatsområde* og dets underbegreber klart fra de andre områder. I Figur 4 er vist fem forskellige grupper af områder, som alle kunne høre under samme inddelingskriterium FORMÅL.

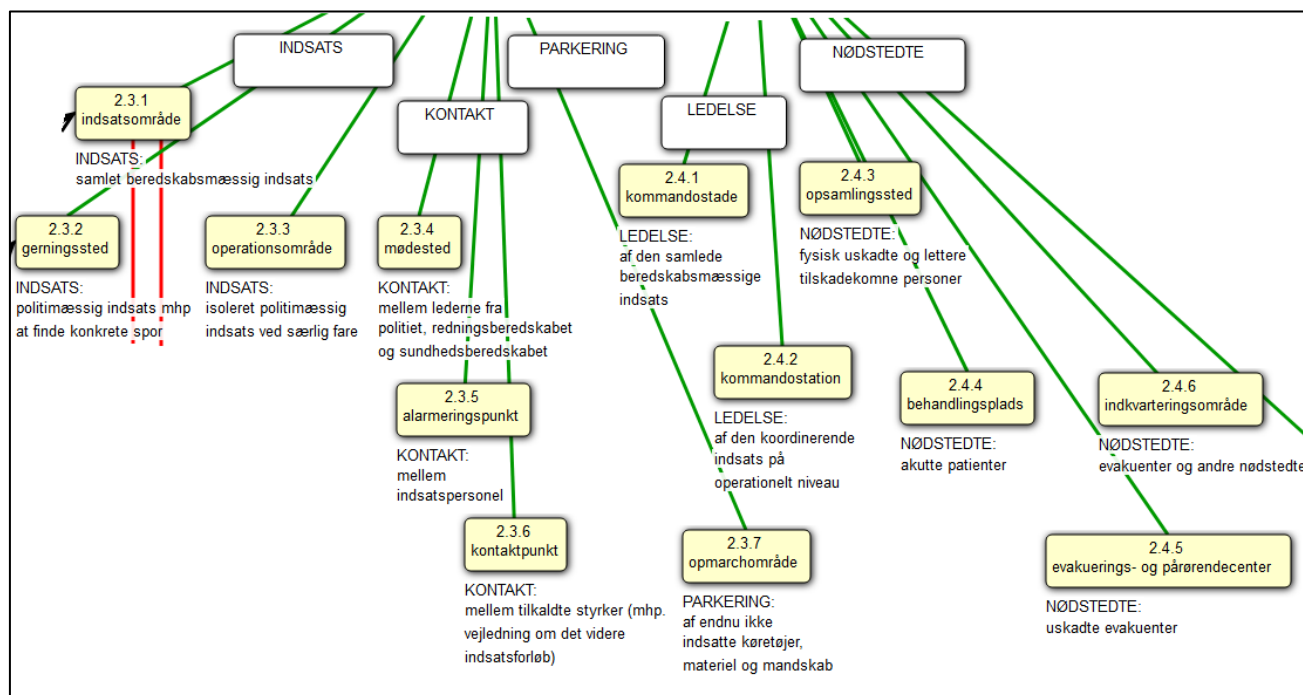
Vi introducerede imidlertid en række mere specifikke inddelingskriterier i stedet for kun FORMÅL, nemlig INDSATS, KONTAKT, PARKERING, LEDELSE og NØDSTEDTE, og fik på den måde opdelt underbegreberne til *område* i flere grupper.

3.2 Inddelingskriterier vs. leksikalisering af ikke-leksikaliserede begreber

En alternativ metode til at opnå underinddelinger ville være at introducere en række ikke-leksikaliserede begreber. Dette ville måske give en klarere inddeling end inddelingen ved hjælp af inddelingskriterier. Ved ikke-leksikaliserede begreber forstår vi begreber, som ikke omtales ved hjælp af en term, men som eksisterer logisk (Damhus et al. 2009: 46-49).

I Figur 5 ses fire ikke-leksikaliserede begreber: *lokation for sikkerhedshændelse*, *lokation for kontakt*, *lokation for ledelse* og *lokation for nødstedte*.

I denne sammenhæng introducerede vi begrebet *lokation* i stedet for begrebet *område*. Det vil sige, at vi ikke længere har begrebet *område* i vores ontologi, jævnfør mere nedenfor.



Figur 4: Forskellige områder under fem forskellige inddelingskriterier

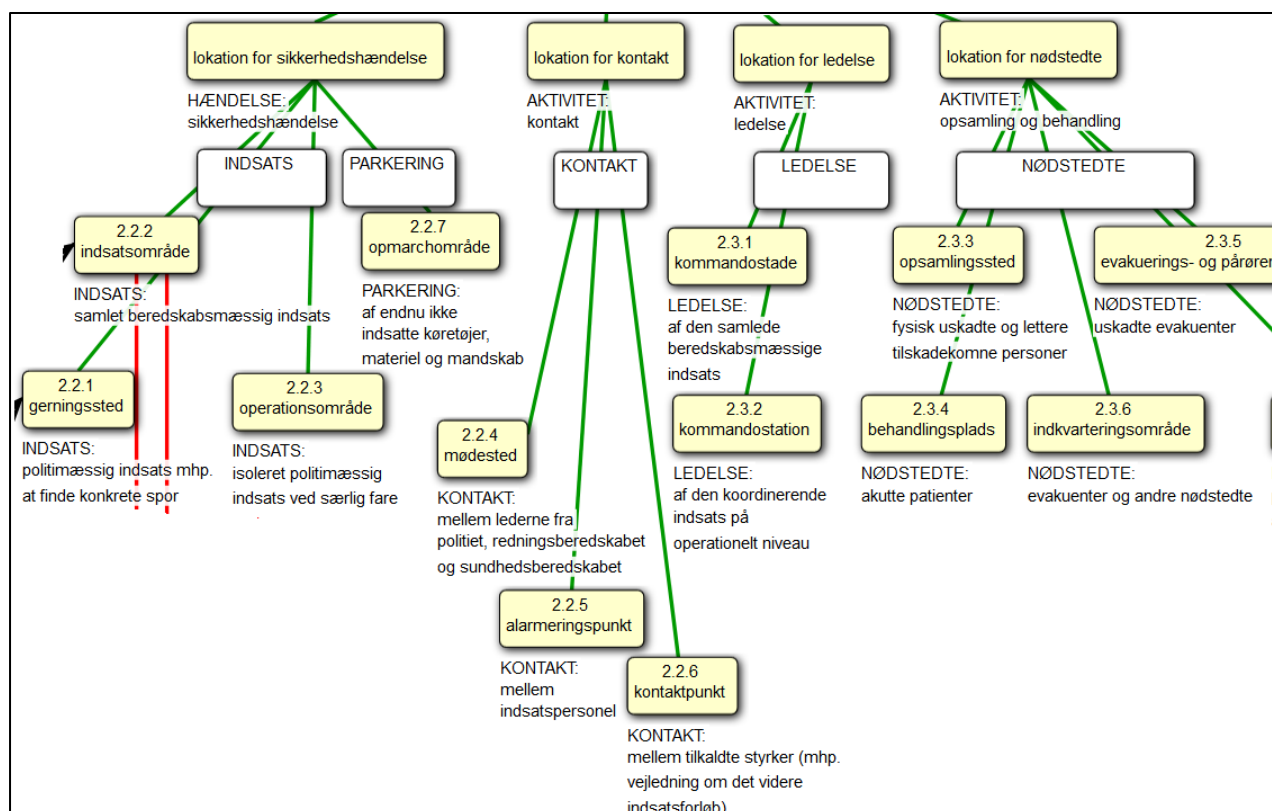
3.3 Problemer med manglende generelle begreber i termlisten

I Beredskabsstyrelsens termliste (2018:64-70) fandtes ikke generelle begreber og definitioner af generelle begreber relateret til sted, for eksempel *område*, *lokation*, *lokalitet*, *aktør* og *aktivitet*. Det betød, at vi, på basis af listen alene, ikke kunne skabe en komplet ontologi. Vi havde en hypotese om, at vi kunne finde nogle af disse begreber i det danske netværk FORVIR's Begrebsbase (<https://forvir.terms.dk/>).

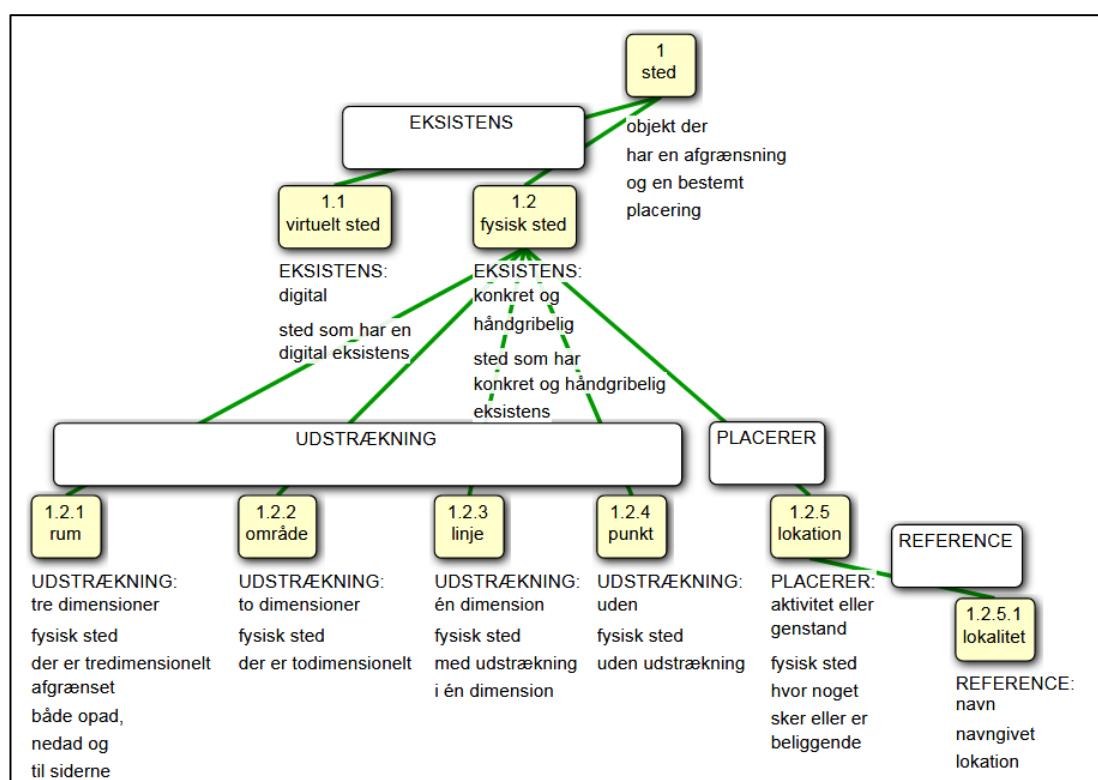
Denne indeholder en ontologi over generelle begreber (en fællesoffentlig topontologi). Derudover antog vi, at vi kunne bruge et diagram vedr. begrebet *sted*, som DANTERMcentret udarbejdede i en rapport til Region Midtjylland, der skulle indrette et nyt supersygehus (DANTERMcentret 2013).

I forbindelse med udarbejdelse af denne rapport opstillede DANTERMcentret et diagram med centrale stedbegreber, hvor de tre begreber *område*, *lokation* og *lokalitet* fandtes, se Figur 6: Udsnit af diagrammet vedrørende *sted* (DANTERMcentret 2013). Vores intention var at 'genbruge' disse generelle begreber i ontologien for indsatsledelse. Vi forsøgte derfor i første omgang at bruge begrebet *område* som overbegreb for *indsatsområde*, *gerningssted* og *operationsområde*. Trækket og definitionen for *område* var i DANTERM (2013):

Karakteristisk træk	UDSTRÆKNING: to dimensioner med geografisk afgrænsning
Definition	fysisk sted der er todimensionalt



Figur 5: Inddeling ved hjælp af ikke-leksikaliserede begreber



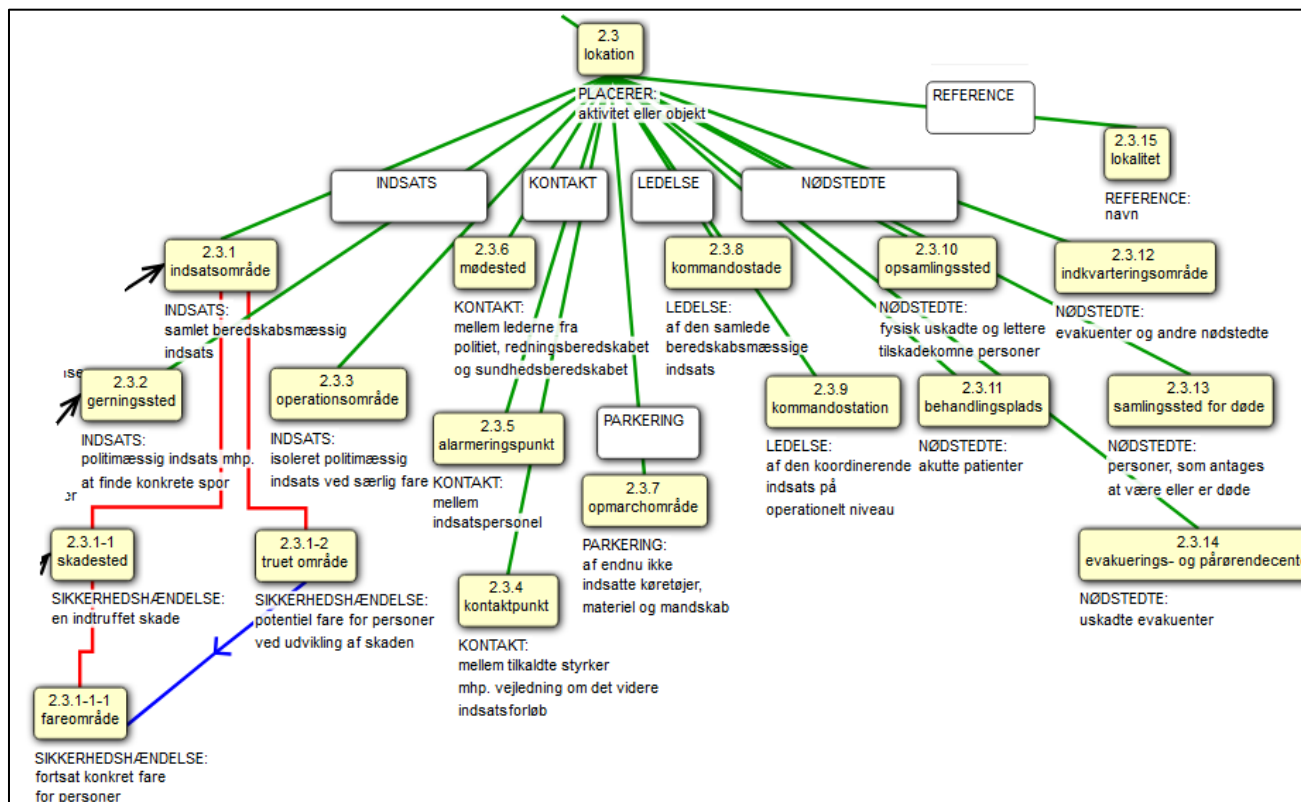
Figur 6: Udsnit af diagrammet vedrørende sted (DANTERMcentret 2013)

Men det var ikke det begreb, vi havde brug for, da dette snarere var noget man kunne kalde en slags *geodata-begreb*. I diagrammet fra rapporten til Region Midtjylland, Figur 6, findes også begreberne *lokation* og *lokalitet* med følgende definitioner:

lokation sted, hvor noget sker eller er beliggende
lokalitet navngiven lokation (for eksempel en bestemt bygning)

Det var netop begrebet *lokation*, vi havde brug for i stedet for *område*, og derfor bruger vi nu det som overbegreb for alle områderne fra termlisten. Som overordnet begreb for *lokation*, *vej* og *afgrænsning* anvender vi begrebet *sted* med definitionen fra (DANTERMcentret 2013), jævnfør Figur 6. Den seneste version af den del af vores terminologiske ontologi, som indeholder underbegreber til *lokation*, er vist i Figur 7.

I den fællesoffentlige ontologi FORVIR (<https://forvir.item.dk/>) findes nogle andre generelle begreber, som vi kunne overtage direkte, for eksempel overbegrebet *aktivitet* og en række underbegreber. Arbejdet med topontologien startede i en arbejdsgruppe under det Nationale Begrebsråd for Sundhedsområdet (under Sundhedsstyrelsen). Afklaringen af forskellene mellem disse begreber var meget tidskrævende, og det har været af stor værdi for os at trække på de to ontologier fra Region Midtjylland og FORVIR.



Figur 7: Den seneste version af en del af den terminologiske ontologi vedrørende *indsatsledelse*

3.4 Omskrivning af forklaringer til intensionale definitioner og træk-værdi par

Endnu et formål med dette indlæg er at redegøre for, hvordan det har været muligt at omskrive forklaringer i termlisten til præcise og konsistente intensionale definitioner og til attribut-værdi-par. Her vil vi tage udgangspunkt i det centrale begreb *lokation* og de sideordnede underbegreber *indsatsområde*, *gerningssted* og *operationsområde*. Den traditionelle intensionale definition er som bekendt den mest præcise måde at definere begreber på. Det sker ved, at definitionerne kun indeholder information om det nærmeste overbegreb og begrebets adskillende karakteristiske træk (ISO 704 (2009:22)).

Hovedformålet med intensionale definitioner er således at adskille det aktuelle begreb fra relaterede begreber (ISO 704 (2009:22); ISO 1087 (2000:6); Erdman Thomsen (2017). En definition skal kun beskrive ét begreb (ISO 704 (2009:28)).

Tabel 1 er et udsnit af en systematisk liste genereret i i-Term. Listen består af de udvalgte begreber, systematiske numre (notationer) fra den terminologiske ontologi, som vi har udarbejdet, og attribut-værdi-par, som vi har identificeret under udarbejdelsen af ontologien. For underbegreberne indeholder oversigten forklaringer fra Beredskabsstyrelsens publikation og de definitioner, som vi har udarbejdet.

De fleste af forklaringerne i termlisten minder om encyklopædiske forklaringer. De indeholder ganske vist termen for nærmeste overbegreb og input til de adskillende karakteristiske træk, men spredt i teksterne, se understregningerne.

Forklaringerne er lange (flere sætninger) og indeholder supplerende informationer, som er overflødige i intensionale definitioner (ISO 704 2009:29).

Begreb	Notation	Attribut-værdi par	Forklaring i termliste	Definition foreslået af os
lokation	2.3	PLACERER: aktivitet eller objekt	Findes ikke i termlisten i Beredskabsstyrelsen (2018)	sted hvor noget sker eller er beliggende (DANTERM 2013)
indsatsområde	2.3.1	INDSATS: samlet beredskabsmæssig indsats	Det samlede <u>område, hvor en beredskabsmæssig indsats foregår</u> . Indsatsområdet er således indsatsledelsens arbejds- og ansvarsområde. Indsatsområdet afgrænses af den ydre afspærring. Se også skadested. Beredskabsstyrelsen (2018:66)	lokation hvor den samlede beredskabsmæssige indsats foregår
gerningssted	2.3.2	INDSATS: politimæssig indsats mhp. at finde konkrete spor	Ud over disse tre områder kan der i <u>politimæssig</u> forstand være tale om et gerningssted. Gerningsstedet vil typisk have en større udstrækning end skadestedet og omfatter hele det <u>område, hvor der er mulighed for at finde konkrete spor</u> . Beredskabsstyrelsen (2018:7)	lokation hvor der foregår en politimæssig indsats for at finde mulige, konkrete spor
operationsområde	2.3.3	INDSATS: isoleret politimæssig indsats ved særlig fare	Et afgrænset <u>område, hvor der foregår eller skal foregå en isoleret politimæssig indsats</u> , som udføres af særligt uddannet politipersonale. Et operationsområde oprettes ved en hændelse, som i indhold og/eller omfang <u>udgør en særlig fare</u> for mennesker eller værdier. Beredskabsstyrelsen (2018:68)	lokation hvor der foregår en isoleret politimæssig indsats ved særlig fare

Tabel 1: Omskrivning af termlistens forklaringer til intensionale definitioner

Da begreberne *indsatsområde*, *gerningssted* og *operationsområde* er sideordnede begreber og alle nærmeste underbegreber til *lokation* i en typerelation, indledes den intensionale definition med termen for det nærmeste overbegreb, det vil sige *lokation*. Det skal dog nævnes, at termen for det nærmeste overbegreb ikke i alle tilfælde kunne uddrages direkte fra termlisten, da der er eksempler på, at der er anført to termer for nærmeste overbegreb, for eksempel:

fareområde: *Det område eller den lokalitet ...*
evakuerings- og *En bygning eller anden*
pårørendecenter: *vejrbestandig lokalitet ...*

Ifølge definitionen i DANTERM (2013) defineres *lokalitet* som en navngiven lokation, det vil sige, at en bestemt *bygning* er et eksempel på en *lokalitet*. *Lokalitet* er også i vores ontologi underbegreb til *lokation*, jævnfør Figur 7. Med hensyn til de adskillende karakteristiske træk, så er alle underbegreberne under *lokation* i Tabel 1 placeret under inddelingskriteriet INDSATS.

Hvad angår beskrivelsen af begrebet *indsatsområde*, så karakteriseres den indsats som foregår i denne *lokation* som beredskabsmæssig. De to sidste sætninger i forklaringen relaterer begrebet til andre begreber i ontologien, nemlig *ydre afspærring* og *skadested*.

Denne information skal ikke indgå i definitionen, men fremgår af ontologien og kan tilføjes i et kommentarfelt i en termbase.

Begrebet *gerningssted* er ikke forklaret i termlisten – og det på trods af, at begrebet omtales i begyndelsen af publikationen. Den manglende forklaring skyldes efter vores overbevisning, at begreberne ikke er beskrevet ud fra en systematisk tilgang, hvor begreberne indbyrdes relationer er undersøgt. I indledningen i Beredskabsstyrelsen (2018:7) findes et tekststykke, som kan anvendes som input til en definition, se tekst og understregninger. Men som det ses, optræder det nærmeste overbegreb og et af de karakteristiske træk først sidst i forklaringen, hvilket forstyrrer forståelsen.

Forklaringen af det sidste begreb *operationsområde* indeholder det karakteristiske træk *afgrænset*, som er overflødigt, da det er nedarvet fra begrebet *sted* (som er overbegreb til *lokation*): *objekt der har en bestemt placering og begrænset størrelse* (DANTERM (2013), se Figur 6. Af ledsætningen i Tabel 1 “hvor der foregår” fremgår det, at der er tale om en *politimæssig indsats*, mens der først i næste sætning og sidst i forklaringen tilføjes endnu et karakteristisk træk, nemlig at der er tale om en lokation (et område), der oprettes ved hændelser, som *udgør en særlig fare*.

I attribut-værdi-parrene svarer attributterne til inddelingskriterier. Om man først udarbejder indholdsdefinitionerne eller attribut-værdi-parrene kan være situationsbestemt eller afhængigt af, hvordan man foretrækker at arbejde. I vores arbejde startede vi med at uddrage attribut-værdi-parrene fra forklaringerne i termlisten, og herefter tilpassede vi dem ud fra de

relaterede begreber i den terminologiske ontologi for *indsatsledelse*.

Forklaringerne, definitionerne og attribut-værdi-parrene i oversigten er nogle af de måder, man kan beskrive begreber på. Valget vil afhænge af formål og målgrupper, om der er tale om mennesker (lægmænd eller eksperter) eller it-systemer. Man kan sige, at der er tale om en stigende formaliseringsgrad, hvor forklaringerne formuleres i fri tekst uden krav til formuleringerne. Herefter følger de intensionale definitioner, som i naturligt sprog formuleres i udsagn ud fra formlen: nærmeste overbegreb og adskillende karakteristisk træk, mens attribut-værdi-parrene formuleres på en formel bestående af ATTRIBUT og VÆRDI. Mens målgruppen for forklaringerne og definitionerne vil være mennesker, vil attribut-værdi-parrene kunne anvendes som grundlaget for datamodellering og dermed it-systemer.

4. Perspektivering

Vi vil til sidst kort perspektivere vores arbejde i forhold til en rapport udarbejdet af Sprogteknologiudvalget under Dansk Sprognævn. Af udvalgets anbefalinger fremgår det, at ”det tekniske og resurse-mæssige grundlag for udvikling af sprogteknologi for det danske sprog ikke [er] på højde med sprogteknologien for hovedsprogene i de lande som vi ofte sammenligner os med (for eksempel Norge, Sverige, Finland). På flere områder ligger dansk blandt de lavest rangerede sprog i Europa” (Dansk Sprognævn 2019:64).

Vi har undersøgt dækningen af de danske termer fra termlisten (inden for indsatsledelse) i et par internationale terminologiske ressourcer, nemlig EU’s terminologibank IATE og tesaurus EuroVoc samt Det Europæiske Miljøagenturs tesaurus GEMET, og har måttet konstatere, at der er ingen eller begrænset dækning.

Behovet for begrebssystemer/terminologiske ontologier og tesaurusser understreges i rapporten fra Sprogteknologiudvalget (Dansk Sprognævn 2019:103). Her nævnes det som resultat af spørgeskemaundersøgelse for terminologidelen: ”Overraskende mange meldte også om behov for en strukturering af begreberne i forhold til hinanden i et begrebssystem, en taksonomi eller en tesaurus. Det er således ikke blot termerne og deres oversættelse der er behov for, men også information om deres indbyrdes sammenhæng”.

Som vi allerede har antydnet ovenfor vil terminologiske ontologier med de formaliserede attribut-værdi-par udgøre et godt grundlag for at opnå enighed om definitioner af begreber og som forarbejde til datamodellering, som vi i Danmark i de senere år har set det inden for det offentlige digitaliseringsarbejde. Tesaurusserne vil kunne anvendes til systemer til automatisk informationssøgning.

Endelig vil vi nævne, at der tidligere har været tværnordisk terminologisamarbejde inden for områder som jobmobilitet, og i Norge arbejdes der med terminologi inden for katastrofehandtering. På Nordterm-konferencen viste det sig, at også den svenske Socialstyrelsen har arbejdet med de begreber, som er behandlet i vores ontologi. Det kunne derfor være interessant at arbejde videre med indsamling af terminologiske ressourcer inden for katastrofehandtering på tværs af de nordiske lande.

5. Referencer

- Aasgaard, Marianne (2017). SPRÅKET I NØDETATER – forklare, ikke forvirre. I: Ágústa Þorbergsdóttir (ed.): *Forvaltning af fagsprog i samfundet, Hvem har ansvaret? Hvem tager ansvaret?* REYKJAVÍK 9.-12. Juni 2015, NORDTERM 19, Island, 83-85.
- Beredskabsstyrelsen (2018). *Retningslinjer for indsatsledelse*. (<https://brs.dk/viden/publikationer/Documents/Retningslinjer%20for%20indsatsledelse%202018.pdf>). (tilgået 1. juli 2019).
- Christensen, Lotte Weilgaard & Bodil Nistrup Madsen (2019). A Danish Terminological Ontology of Incident Management in the Field of Disaster Management (under udarbejdelse)
- Damhus, Ture, Peder Olesen Larsen, Bodil Nistrup Madsen og Signe Zambach (2009). Begrebsafklaring inden for området enzymkemi. I: *Nordterm 16, Proceedings from NORDTERM 2009, Ontologier og taksonomier*, Copenhagen Business School, 10.-11. juni 2009.
- Dansk Sprognævn (2019). *Sprogteknologiudvalget i Danmark* (<http://sprogtek2018.dk>) (tilgået 1. juli 2019).
- DANTERMcentret (2013). *Begrebsmodellering af stedbegrebet i hospitalssammenhæng*. (<https://www.dnu.rm.dk/siteassets/dnudnu/it/sporbarhed/begrebsmodellering-af-stedbegrebet-i-hospitalssammenhang-rapport.pdf>) (tilgået 1. juli 2019).
- EuroVoc. *EU Vocabularies* (<http://eurovoc.europa.eu/>) (tilgået 1. juli 2019).
- FORVIR. *Forum for vidensmodellering i offentligt regi* (<https://forvir.iterm.dk/>) (tilgået 27. juni 2019).
- GEMET. *General Multilingual Environmental Thesaurus* (<https://www.eionet.europa.eu/gemet/en/about/>), (<https://www.eionet.europa.eu/gemet/en/groups/>) (tilgået 1. juli 2019).
- IATE. *InterActive Terminology for Europe* (<http://iate.europa.eu/>) (tilgået 1. juli 2019).
- i-Term. Terminology and Knowledge Management System, DANTERM Technologies, <https://www.danterm.dk/> (tilgået 15. august 2019)
- ISO 704 (2009) *Terminology work – Principles and methods*. Genève: ISO.
- ISO 1087-1 (2000) *Terminology Work – Vocabulary – Part 1: Theory and Application*. Genève: ISO.
- Rosenfeld, Louis, Peter Morville & Jorge Arango (2015): *Information Architecture: for the Web and Beyond*. Sebastopol, California: O’Reilly Media.
- Spencer, Donna (2009): *Card sorting: Designing usable categories*. Brooklyn, New York: Rosenfeld Media.
- Stoyanov, Stoyan (2017). Crisis and Disaster Management Terminology. In: Orlin Nikolov and Swathi Veeravalli (eds.): *Implications of Climate Change and Disasters on Military Activities*, NATO Science for Peace and Security Series C: Environmental Security. https://link.springer.com/chapter/10.1007/978-94-024-1071-6_2 (tilgået 1. juli 2019).
- Thomsen, Hanne Erdman (2017). The intensional definition. In: Rute Costa and Christophe Roche (eds.): *The definition in Terminology*, Berlin: Frank & Timme, Verlag für wissenschaftliche Literatur. (skal verificeres)
- UNISDR (2016). *UNISDR Terminology on Disaster Risk Reduction*. United Nations International Strategy for Disaster Reduction, Geneva, 35 pp.

Språkressurser i språkteknologien – termenes uerstattelighet og hvordan overtale offentlige etater til å ta ansvar for språkteknologi

Kristine Eide

Språkrådet

kristine.eide@sprakradet.no

I digitaliseringsprosessene i flere skandinaviske land, er begrepsarbeid en viktig premiss for at offentlige etater skal kommunisere med hverandre, og flerspråklige termlister er nødvendige når kommunikasjonen foregår over landegrenser. Offentlige etater må dermed ta ansvar for begrepsapparatet innen deres område.

Resultatet av dette begrepsarbeidet kan gjenbrukes til andre formål. En av de mest grunnleggende ressursene for automatisk oversettelse og mye annen språkteknologi er nettopp termlister.

Moderne språkteknologi fungerer foreløpig best på avgrensede fagområder. For at en talegjenkjenner skal forstå hva legene sier, må den føres med medisinske

uttrykk. En banks praterobot, må føres med bankterminologi. Jo mer spissede data en automatisk oversettelse arbeider med, jo bedre blir oversettelsen innenfor det samme avgrensede område. Derfor er for eksempel EUs system for automatisk oversettelse laget slik at man kan spesifisere innenfor hvilket område en tekst skal oversettes.

En termliste som er utarbeidet for et spesifikt område, kan dermed gjenbrukes til andre formål enn den opprinnelig var tiltenkt, vel og merke hvis den er tilrettelagt for gjenbruk. Dette må tas hensyn til i begrepsarbeid, og særlig offentlige etater må være klar over hvilken merverdi resultatet av terminologiarbeid kan tilføre både egen etat og samfunnet for øvrig.

A Practical Example of the Use of Terms in Automated Classification

Brian Jacobsen

Chefkonsulent

Taxon ApS, Østergade 1, 1100 København K

Brian@Taxon.dk, +45 2390 5404

Keywords: Automated Classification, Process Automation, Metadata

An important part of automating processes is the ability to classify (or “tag”) texts according to one or more taxonomies.

The resulting metadata is vital in order to exchange, find and verify data between various systems, internal as well as external.

Terms in the sense of one or more coherent words are the very foundation of the automated classification system.

The presentation presents the automated classification system Taxon Classifier and the corresponding taxonomy tool TaxonHub. Together the two systems facilitates automated classification.

This will include a discussion about terms that combined defines a domain specific class, i.e. “super synonyms”, the use of statistics in combination with a rule engine in automated classification as well as practical examples of running systems like automated distribution of incoming mails using OS2KLE (the taxonomy used by all Danish municipalities) and access control in ESDH/ECM.

The goal of the presentation is to pave the way to a greater understanding of the important role terminology plays in automated classification and the automatization of processes in general.

Nordic European Terminology Collaboration and Networking

Stine Jensen

Terminologist, TermCoord, European Parliament
Plt. du Kirchberg, L-2929 Luxembourg
dgtrad.termcoord@europarl.europa.eu
/stinekyungjensen@gmail.com
+352 4300 23872
/+45 61 26 21 60

Abstract

Terminology networking, expert involvement and collaboration is not a new thing within EU terminology, and different language communities have already established language councils and networks for such purpose. For example: Danish (the Danish Terminology Network), Irish (the Expert Question Box), and Italian (REI - Network for the Excellence of Institutional Italian). Building upon this, the Terminology Coordination Unit of the European Parliament (TermCoord) last year made an effort to enforce this kind of networking with the further development of the [Expertise Network](#).

The objective of this collaborative terminology is to have the best possible quality in the database for European Terminology IATE.

Since 2013, TermCoord has formalised terminology collaboration with terminology or language departments of universities and other interested parties: such as Translators Without Borders on terminology projects. A recent example of such collaboration is [Terminology without Borders](#), which combines expert involvement, expert validation and university collaboration within different fields: Health, fisheries, Education and women's rights.

The participating universities are predominantly from Southern European countries. As a result, TermCoord has, for the last year been working on balancing the representation of networks by strengthening the Nordic representation in collaboration with the Nordic language communities, such as Danish, Finnish and Latvian.

Collaborative Terminology and Language Technology

Terminology networking, expert involvement and collaboration is not a new thing within EU terminology, and different language communities have already established language councils and networks for such purpose. For example: Danish (the Danish Terminology Network), Irish (the Expert Question Box), and Italian (REI - Network for the Excellence of Institutional Italian).



Figure 1: Terminology Councils and Networks¹

The coordination of such terminology collaboration takes place at two different levels which can be inter- (between European Institutions) or *intra*institutional (within the same European Institution):

- 1) Language independent and central terminology coordination for the language communities, representing the official EU languages², i.e. TermCoord
- 2) Language specific and interinstitutional terminology coordination, i.e. the Danish language community.

On the first level, the central terminology coordination takes place within one or several of the European Institutions, such as the European Parliament, the European Council and the European Commission. A collaboration does not necessarily include all of them.

The terminology collaboration at this level is normally language *independent* with the aim of serving all official EU languages.

In this context, the collaborative nature of the Discharge terminology carried out between the thematic group³ for the EU Budget, the Budgetary Group, in the Danish Translation Unit of the European Parliament and TermCoord last year to

¹ (Language Wikis, 2019).

² EU official languages: Bulgarian, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, German, Greek, Hungarian, Irish, Italian, Latvian,

Lithuanian, Maltese, Polish, Portuguese, Romanian, Slovak, Slovenian, Spanish, Swedish.

³ The EU Translation is organised in thematic groups according to the translation needs.

create a new Discharge Collection⁴ can be mentioned. This collaboration involved several both national and European Institutions, such as the banks and the Court of Auditors and the European Commission.

The second level is language *specific* and can be within the same or several European Institutions, meaning that this type of terminology collaboration comes from the language communities themselves, either within one or several language communities.

A good example of this is the expert list of the Swedish language community which together with the Terminology Council for the Low Countries was the inspiration source for the establishment of the Danish Terminology Network (*Termnetværket EU-DK*). The last one as a result of Frieda Steur's intervention at the VIII European Terminology Summit "Visions and Revisions" hosted by TermCoord in Luxembourg in 2016 (Jensen, 2018).

Having said that, it is important to note that the levels do not necessarily reflect the order of origin of a given terminology collaboration.

Collaboration or terminology initiatives often originate from language level as in the case of the Danish Terminology Network, the Irish Expert Question Box and the Italian Network for the Excellence of Institutional Italian.

These collaborations were the motive behind TermCoord's effort last year to enforce this kind of networking with the further development of the Expertise Network⁵ by including all in-house expertise from the agencies and EU related bodies.

Expertise Network

If you need to consult a subject matter expert, here you can find who can help you with your translation.

You can filter this list by language, by domain, and then just by clicking on a contact name you can send an email to the expert.

DOMAIN = Click and start type LANGUAGE = Click and start type ORGANISATION = Click and start type

Hide Columns = COUNTRY X

	DOMAIN	LANGUAGE	ORGANISATION	URL/EMAIL
1	Safety and Health at work	BG	EU-OSHA National Focal Points	https://osha.europa.eu/en

Figure 2: Expertise Network⁶

The promotion of best practice of some language communities to inspire others to follow is one of TermCoord's finest assignments.

⁴ Collections are subject specific terminology collections, created for a specific purpose. In this case, it was Discharge terminology which the thematic group of the EU Budget, the Budgetary Group, needed for the Discharge translations. The terminology collections are downloadable from IATE.

⁵ The Expertise Network is language specific but serves all EU languages and is made by the Central Terminology

Nordic Representation

Since 2012, TermCoord has *formalised* terminology collaboration with terminology or language departments of universities on specific terminology projects. The representation is, however, to some extent reflecting the existing interest in languages, translation and terminology in the given mother country⁷ as collaboration has occurred mostly according to the interest of the collaboration partner which, as you can see from figure 3, does not point North.



Figure 3: Terminology Collaboration⁸

This is unfortunate and might endanger linguistic equality and representation, which is important for the development of EU language technology, and which is consolidated in Article 22 'Cultural, religious and linguistic diversity' of the Charter of Fundamental Rights: 'The Union shall respect cultural, religious and linguistic diversity (EU, 2012).'

The EU understanding of multilingualism has no precedent (EU, 2019) and is the most 'complete' in the Translation Business as EU law enshrines the principle of equal authenticity (principle of language equality) whereby all language versions of EU legislation are authentic and are presumed to have the same meaning (European Parliament, 2017).

This means that all translations, once in force, are considered originals and the aim of EU multilingualism is to speak with one voice (Jensen, 2017).

The consequences of linguistic inequality might eventually affect the quality and recompilation of the language resources used to feed the EU language

Coordination in the European Parliament as an offer to all language communities across EU Institutions.

⁶ (EURTerm, 2019).

⁷ Mother country meaning the country of the language community, i.e. Finland for Finnish, Sweden for Swedish, etc.

⁸ (Jensen, Stine, presentation Nordterm 2019; Terminology Coordination, 2019).

technology, as for example eTranslation, Euramis⁹ (European Commission, 2019), IATE, the Language Wikis, etc.

Such consequences spread to the national level and to some extent international level as the language resources from big public organisations with considerable translation and interpretation services, i.e. the EU and the UN, are often used for feeding national and international language technology. This was the case with Google Translate (Statistics Views, 2015).

AI.Tech

In relation to the subject of Nordterm 2019, the Role of Terminology in Language Technological Applications, AI.Tech, the new section on TermCoord's webpage termcoord.eu was presented.

AI.Tech

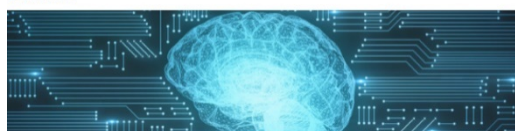


Figure 4: AI.Tech¹⁰

It is dedicated to the subject of language technology and terminology in the era of Artificial Intelligence (AI) and inspired by the latest conferences on the matter to which TermCoord has attended, such as JIAMCATT 2019.

In the article of Rodolfo Maslias, Head of the Terminology Coordination Unit in the European Parliament, 'Terminology in the Brain of the Machine', the initiatives and contributions of TermCoord to prepare and improve the terminology resources are lined up:

'We realise that NMT¹¹ can be forced to use only very high-quality terminology resources, i.e. highly reliable, normative terms. This led us to step up our efforts to eliminate the noise from our database, to clean it from old and outdated data and to create domain-related collections offering advanced filtering possibilities. Facing this new challenge for terminology, we also have to reconsider our policy consisting in inserting our resources and features in the pre-translation phase and to focus more on the post-editing and quality checking tools (Terminology Coordination, 2019).'

The EU Context

Apart from the representation and linguistic equality, the EU context makes terminology

networking, expert involvement and collaboration extremely important due to three factors:

The first one being that such collaboration is *encouraged* but not obliged and is as it is now built upon voluntary action both from central coordination and from the language communities.

This purpose is served, for example, by the IATE Management Group (the steering group of IATE) and the working groups installed by it, where a huge effort is made to avoid double work. The new [IATE](http://iate.europa.eu) is an example of such effort.

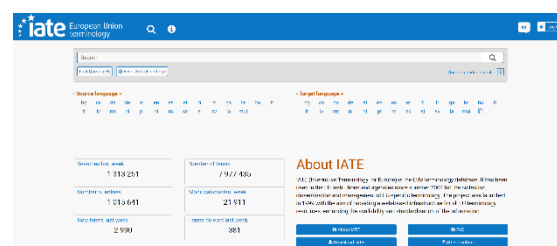


Figure 5: The New IATE¹²

However, there is so far no 'central coordination of the central coordination'. Instead, each European Institution has their own central terminology coordination and these central coordinations all have to work together.

The second one being the physical distance between the translation services and the mother country and the influence from other languages and cultures that the single mediator¹³ might experience as languages are dynamic, non-static and therefore change over time.

This dynamic character of languages is inevitable and goes hand in hand with the change of context, of the reality of people, as for example social media and/or the power balance in the World, i.e. currently English and earlier French and German. Such change of reality will eventually start reflecting itself into the language.

The reason is that we need new words to describe what we do and what we live, which might also explain why new phenomena that appear in one language, stays in that language and then might become nationalised later on.

To stay with the example of social media, the verb 'tweeting' comes from the social media, Twitter (Twitter, 2019), and the expression of something being 'insta-friendly' means that it is suitable for a

⁹ Euramis (European advanced multilingual information system) is the shared translation memory of the European Institutions (European Commission, 2019).

¹⁰ (TermCoord, 2019).

¹¹ NMT: Neural Machine Translation.

¹² (IATE, 2019).

¹³ Mediator here includes both translators and interpreters.

post on the social media, Instagram (Instagram, 2019). In the same way, the slight change of meaning behind a heart or (a) thump(s) up of 'I like it' – 'it' being a post on social media – rather than 'I like or love *you*', 'good job', etc., comes from the 'Like' (button) on social media (Facebook, 2019; Instagram, 2019; Twitter, 2019).

Even though this example is taken from general language, linguistic dynamics also concerns professional language although it might be argued that this is to a somewhat smaller extent. However, such dynamic happens when, i.e. a new technology, engine or product is created, such as Self-Driving Vehicles, speech recognition, AI, etc.

In terminology, you also talk about the 'creation of new terms'.

This leads us to the third factor which is the 'limitations of human capability' or 'the limitations of human memory' meaning the (reasonable) fact that the EU mediators are mostly generalists and (can)not (be) subject matter expert in all areas of activity of the EU.

As it is now, it is up to each mediator to keep up with linguistic developments, whether belonging to general or professional language, in the mother country *if* a terminology collaboration has not been formalised and put in place.

Contact with subject matter experts, creators and users of the terminology sought within a specific subject area and beyond the Translation Business is therefore important for the EU Translation as a whole in order to make sure that the terminology inserted into IATE and thus used in the EU translations is in line with the current terminology use in the mother country as well as to ensure linguistic equality in the EU language resources and technology.

New type of collaboration

The traditional type of terminology collaboration on specific terminology projects based mostly on external interest has developed into being more than academic and reaching out to a broader audience than the subject matter experts: the terminology users and other interested parties, such as Translators Without Borders (Terminology Coordination, 2019). This new type of collaboration was also presented.

That way every step from the moment of its creation, terminology and its use to *communicate* within different layers of the terminology users, is covered and included in the collaboration:

- 1) *In which terms* do experts communicate with each other?

- 2) In which terms do experts communicate with laymen, including children?

- 3) And vice versa.

An example of such terminology collaboration is the innovative programme of collaborative terminology, [Terminology Without Borders](#):

'Terminology without Borders yourterm.org is a programme run by the Terminology Coordination Unit of the European Parliament that coordinates terminology work between specialised knowledge agents and participating universities with expertise in terminological research (Terminology Coordination, 2019).'

The programme has its origins in the terminology project, [YourTerm Med](#), and includes a specific collaboration with the University of Diderot (Paris) on children's terminology or the terminology used to talk to children within the medical field (Terminology Coordination, 2019).

Terminology Without Borders is innovative, first of all, because TermCoord actively reaches out to potential collaboration partners, both in-house and external.

Second of all, this new approach to terminology used by all terminology users, is integrating the aspect of *communication* and the communication situation using general language and/or terminology *adapted* to the receiver of the message.

This is a much more complex approach than the one we have seen until now where terminology has traditionally been connected to professional language, but it goes hand in hand with two different tendencies within the Translation Business:

The first one being to look at translation as a mediation process, entailing all aspects of communication, which is also reflected in academia, as an example the perception of the University of Salamanca, 'Translation and Intercultural Mediation', can be mentioned (USAL, 2019).

The second one being the new development within terminology of an approximation between terminology or professional language and general language by, i.e. gathering all language resources in one place, such as the [PMKI](#) (European Commission, 2016; Jensen, 2019) and the [Eurotermbank](#) (Eurotermbank, 2019; Nordterm 2019).

Nowadays, we have even begun to talk about 'every-day terminology' or 'plain terminology' (Terminology Coordination, 2019).

Results of Nordterm

TermCoord came to Nordterm as part of last year's intensified work on balancing the representation of networks by strengthening the Nordic representation to offer collaboration as well as interviews with prominent Nordic terminologists.

The idea behind the interviews was to raise awareness, both nationally and inhouse:

- 1) The interviews would serve as an in-house 'teaser' for the EU language communities, representing the Nordic languages, such as Danish, Finnish and Latvian about Nordic Terminology and activity to initiate terminology collaboration
- 2) The interviews should inform about the importance and activity of terminology in general, both in-house and nationally; especially with regard to the interest of the Nordic countries for languages, terminology and translation
- 3) The interviews should inform nationally (the Nordic countries and/or terminologists) about the possibility to collaborate with the European Institutions on the terminology used in for example European legislation, which has direct effect on national legislation.

With us, we brought back some interviews from both academia and the private sector in Denmark. The interviews will be published in TermCoord's webpage termcoord.eu under the section, 'Why is Terminology Your Passion?' (Terminology Coordination, 2019).

Another unexpected and *informal* collaboration request that we received at Nordterm was far bigger than we would have expected and came from outside the EU, from Norway; to add Norwegian to IATE.

This request has since then been presented to the European Institutions at an IATE Management Meeting. During this meeting, Icelandic was also considered for two different reasons:

- 1) Norwegian and Icelandic are both part of the eTranslation project and Euramis (European Language Resource Coordination, 2019)
- 2) The special status of Icelandic and Norwegian due to the EEA agreement (EAFT, 2015).

A formalisation of this request should be on its way.

Bibliographical References

Centre for Language Technology – Department of Nordic Studies and Linguistics – University of Copenhagen. (2019). Nordterm 2019,

- <https://cst.ku.dk/arrangementer/nordterm-2019/> (1.10.2019)
- EFTA. (8.4.2015). Improved access to Icelandic and Norwegian translations of EU legal acts, <https://www.efta.int/EEA/news/Improved-access-Icelandic-and-Norwegian-translations-EU-legal-acts-3891> (25.9.2019)
- European Commission. About multilingualism policy, https://ec.europa.eu/education/policies/multilingualism/about-multilingualism-policy_en (26.9.2019)
- European Commission. (11.4.2019). DGT- Translation Memory, <https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory> (26.9.2019)
- European Commission. (2016). Overcoming Language barriers, https://ec.europa.eu/isa2/actions/overcoming-language-barriers_en (25.9.2019)
- European Commission. The Charter of Fundamental Rights, https://ec.europa.eu/info/aid-development-cooperation-fundamental-rights/your-rights-eu/eu-charter-fundamental-rights_en (24.9.2019)
- European Language Resource Coordination, Frequently Asked Questions – and their Answers, <http://www.lr-coordination.eu/faq> (25.9.2019)
- European Parliament. (January 2017). Legal aspects of EU multilingualism, http://www.europarl.europa.eu/RegData/etudes/BRIE/2017/595914/EPRS_BRI%282017%29595914_EN.pdf (26.9.2019)
- European Parliament, Think Thank. (26.1.2017). Legal aspects of EU multilingualism, [http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI\(2017\)595914](http://www.europarl.europa.eu/thinktank/en/document.html?reference=EPRS_BRI(2017)595914) (24.9.2019)
- EU. EU Languages, https://europa.eu/european-union/about-eu/eu-languages_en (26.9.2019)
- EU. (26.10.2012). The Charter of Fundamental Rights of the European Union, in: the Official Journal of the European Union, http://data.europa.eu/eli/treaty/char_2012/oj (24.9.2019)
- Facebook, <https://www.facebook.com/> (1.10.2019)
- Instagram, <https://www.instagram.com/> (1.10.2019)
- Jensen, Stine. (31.5.2018). EP Networking with Denmark for Terminology, <https://termcoord.eu/2018/05/ep-networking-with-denmark-for-terminology/> (24.9.2019)
- Jensen, Stine. (20.6.2019). Semantic interoperability for the multilingual web, <https://termcoord.eu/2019/06/semantic-interoperability-for-the-multilingual-web/> (25.9.2019)
- Jensen, Stine. (23.10.2017). Terminology and Multilingualism, <https://termcoord.eu/2017/10/terminology-and-multilingualism/> (26.9.2019)

JIAMCATT. (2019),
<https://www.iamladp.org/content/jiamcatt>
 (26.9.2019)

Statistics Views. (23.6.2015). The statistics behind
<https://www.statisticsviews.com/details/feature/8065581/The-statistics-behind-Google-Translate.html> (26.9.2019)

Terminology Coordination. AITech,
<https://termcoord.eu/category/aite/> (26.9.2019)

Terminology Coordination. (14.3.2019). Plain Terminology,
<https://termcoord.eu/2019/03/plain-terminology/>
 (25.9.2019)

Terminology Coordination. (16.5.2019). Terminology in the Brain of the Machine,
<https://termcoord.eu/2019/05/terminology-in-the-brain-of-the-machine/> (26.9.2019)

Terminology Coordination. Terminology Without Borders, <https://yourterm.org/> (24.9.2019)

Terminology Coordination. University projects.
<https://termcoord.eu/universities-projects/>
 (24.9.2019)

Terminology Coordination. Why is Terminology your Passion?
<https://termcoord.eu/terminology/why-is-terminology-your-passion/> (25.9.2019)

Twitter, <https://twitter.com/> (1.10.2019)

University of Salamanca. Máster Universitario en Traducción y Mediación Intercultural,
<https://www.usal.es/master-traduccion-mediacion-intercultural> (25.9.2019)

Language Resource References

EU. Language Wikis,
<https://webgate.ec.europa.eu/fpfis/wikis/pages/viewpage.action?pageId=193339056> (30.6.2019)

Eurotermbank,
<https://www.eurotermbank.com/> (27.9.2019)

EURTerm,
https://ec.europa.eu/isa2/actions/overcoming-language-barriers_en (27.9.2019)

Google Translate, <https://translate.google.com/>
 (26.9.2019).

IATE, European Union Terminology,
<https://iate.europa.eu/home> (24.9.2019)

Digitaliserte tekster for terminologisk analyse

Lars Bagøien Johnsen
Det Norske Nasjonalbibliotek
Lars.johnsen@nb.no

Abstract

Nasjonalbiblioteket i Norge har digitalisert hele bokarven, som er gjort tilgjengelig for terminologer og leksikografer. Vi skal spesielt se på hvordan sammenhenger og betydninger kan analyseres ved hjelp av kollokasjoner, og samforekomster i koordineringsmønstre.

Keywords: grafer, nettverk, betydning, cluster

I denne presentasjon vil vi gå gjennom den dataskinnelle tilgangen til tekster ved Nasjonalbiblioteket i Norge. Til forskjell fra den konsumerbare lesning av tekster, som bare er tilgjengelig for norske IP-adresser, er den programmatisk tilgjengelig for alle. Den tilbyr konkordanser, n-grammer, ordgalakser, kollokasjoner og konkordanser.

Dataene som er gjort tilgjengelig for leksikalske undersøkelser, utgjør ca 500 000 digitale bøker, og 800 000 aviser. For bøkene er alt av metadata tilgjengelig, som gjør det mulig å bygge korpus og undersøke ord innenfor bestemte grupper av tekster klassifisert i henhold til Dewey desimalklassifikasjon.

Bruken av metadata gjør det mulig å hente ut konkordanser, og hele setninger som belegg for bestemte betydninger.

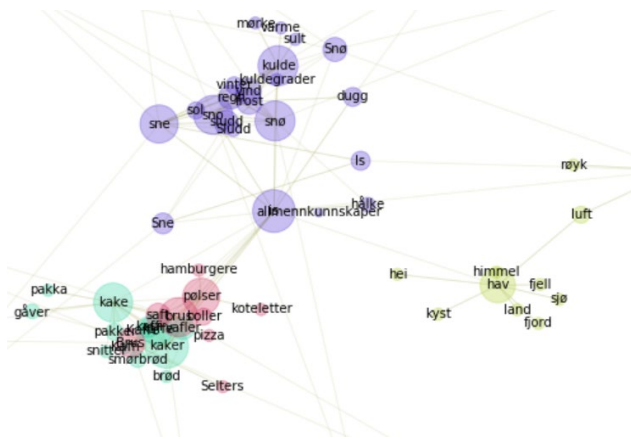
For en leksikograf er dataene tilgjengelig i nettlesere, og benytter Jupyter Notebook som plattform.

Vi har data n-grammer, plotting over tid, og hvordan finne ut hvilke ord som har kraftige brudd i frekvens.

Hvordan henger ord sammen, og hvordan skille de forskjellige mønstrene fra hverandre.

Språkressurser

Dataene er tilgjengelig fra den norske språkbanken, del av Nasjonalbiblioteket.



Figur 1 Koordinasjonsgraf for ordet 'is'

Tekstene er gjort tilgjengelig på en slik måte at man kan analysere ord og deres betydninger gjennom kollokasjoner eller koordinasjonsgrafer. I Figur 1 er en koordinasjonsgraf for ordet 'is' som viser ord som kluster seg sammen. Grafen skiller mellom flere betydninger, der de blå nodene utgjøre det stofflige, som i « is og snø », og de grønne « is » som dessert eller etterrett.

Textile Terminology in the Thread project – a multilingual approach

Susanne Lervad, Christian Gaubert
Visiting scholars at The Centre for Textile Research
Københavns universitet
Karen Blixens Plads 8
DK-2300 København S
sl@termplus.dk, christian.gaubert@laposte.net

Abstract

Presentation of the Thread project at the Centre for textile Research (CTR) at the University of Copenhagen – a 3 year Innovation Fund Project in order to create a model for Education, Empowerment and Entrepreneurship for refugee women in Denmark. The presentation reports on an internship of refugee women from Eritrea and Turkey at CTR in 2017-18, and the experiences and the vocabulary result of the multilingual sessions in Arabic, Danish, English, and Tigrinya on textile terminology. The terminology work includes verbal and non-verbal representation of the concepts such as dress by photos, drawings, and models in order to combine different ways of representing the basic concepts of the field. The report then focuses on the case of textile terminology in Tigrinya, with an overview of the visible influences on textile terminology in this Ethio-Semitic language.

Keywords: terminology, lexicography, textiles, Tigrinya, diachrony, refugees

1. The THREAD project

THREAD is a dynamic and innovative collaboration of diverse partners from business, design, education, research, and refugee support agencies aiming to achieve life-changing results for women who are forging new lives in Denmark.

Integration is a key problem for refugees and their host countries. This project seeks to help address this problem by developing and testing a pioneering themed model of Empowerment, Employment, and Entrepreneurship opportunities. The model offers a step-change programme to women refugees with differing level of life experience and educational qualifications, which may be accessed according to their individual confidence and competence.

2. An internship of Eritrean refugees in CTR

2.1 Involving refugees in terminological researches

In 2017, two refugee women from Eritrea, Alem and Mlete, had internships at the Centre for Textile Research (CTR) for 16 weeks in the frame of THREAD. As the textile terminology expert at CTR, Susanne Lervad devised multilingual sessions and documentation of vocabulary for knitting, tools, and materials in Arabic, Danish, English, and Tigrinya thereby establishing a relationship with the refugees by exchanging knowledge on textile history.

We established a relationship with the two Eritrean refugees by focusing vocabulary work on Tigrinya, which is not a well-documented language especially for textile terms. The two women were native speakers of Eritrean with little understanding of Arabic, for which we had more in-house support. The Centre for Textile Research hosts a website reporting issues about textile craft (www.traditionaltextilecraft.dk) with a section on multilingual terminology as well as different concepts/terms for materials and techniques such as

spinning, weaving, tools (pins and shuttles), and other equipment.

The terminology table had unfinished columns for some Arabic terms. However, it became apparent that the women's own language of Tigrinya also deserved an in-depth study. During the first session, we discussed the tools and materials, such as wool and silk. We documented material use, spinning, and yarn vocabulary in Tigrinya and we started to knit, which the refugees had not learnt before.

The knitting was very beneficial for the refugees' well-being as it is both contemplative and provides a sense of accomplishment. The knitting processes were documented in instructions which included the categories from the previous sessions about material, and spun yarn (fig. 1). The wool and silk yarns that were used for knitting a traditional, triangular Danish shawl were pleasant to work with – soft and warm in a cold Danish winter. Hours passed as we talked about our lives, families, home countries and language but also textile-related issues: Who taught us how to knit? Did our children know the techniques? What are the different knitting techniques that can be used to achieve the same result?

Instruction for a knitted shawl

Material: Silk and wool yarn (thin)

Tool : Pins number 3

Process: Make two loops and knit the first round. Increase with one knit stitch by the end of each row. Continue until the shawl is long enough for the individual person.

Opskrift på strikket sjal

Materiale: Tyndt Silke- og uldgarn (tyndt)

Redskab: Pinde nr 3

Fremgangsmåde: Slå 2 masker op og strik første pind i ret. tag en maske ud i slutningen af hver pind i ret. Fortsæt indtil sjallet har den ønskede størrelse.

መምርሒ ማጽሐፍ ለጽሑፍ ማርባ

እቲ መሰርሒ ከልተ ናይ እጃቲ ባትሪ ቁጽሪ ስለስተ .

ከገሰርሒ ከለና ብ ከልተ ዓይነት ፈትሒ .

ፎክሎር ኣብ ሓገቲ ዙር ሓገቲ እንዳ ወሰኽና ንኺይድ .

ስጋብ እቲ ዝተመሰና ልክዕ ንጅጽል .

Figure 1. Basic instructions for knitting Danish/TN

The instructions were simple. We started with three loops and then increased the number of loops until the shawl was finished. Hand knitting the shawl, and handwriting the terminology was a very simple way to work. The discussion helped us connect through an understanding of dress and how refugees' work with their clothes to adapt to their new life in Denmark. Basic textile vocabulary was gathered (fig. 2) and recorded in sound files at this stage.

English	Loose translit.	Transliteration	Tigrinya
warp	tukul ; komet	tuxul ; komät	ቱኪል ; ቅመት
weft	gadem	gadama	ጋድም
wool	tsogeri nay ensesa	şagari naya 'anasasa	ጸግሪ ናይ እንስሳ
sheep	begi'	bägi'a	በጊዕ
goat	tel	jela	ጤል
to spin	mefhasi	māfahasi	መፍሓስ
yarn	meftel	māfatāla	መፍተል
fiber	fetli	fatali	ፈትሊ
product/result	wetseit serah	wuşıta saraḥa	ወደሊት ስራሕ
process	mesrh	māsaraḥa	መስርሕ
material ; textile	cherki	çaraki	ጨርቂ
knitting	echeneto	'āēaneto	እቕኔቶ

Figure 2. Sample of tigrinian textile vocabulary

When it came to the evaluation of the internships, Alem and Mlete's comments were very strong and positive about these word/practice sessions. They prepared a Powerpoint presentation of their experiences and gave a short talk explaining some of the Eritrean dress traditions. They also described their work with knitting terminology. Today, contact has continued between the academics at CTR and the refugees via social media, and social engagements in museums as well as their homes. We as academics have now presented the Tigrinya vocabulary at several terminology conferences, and in terminological newsletters with acknowledgment of Alem and Mlete's contributions.

2.2 Eritrean refugees in Denmark

The case of Eritrean refugees is relevant at the moment in Denmark as the top three refugee nationalities reported in 2018 were Eritrean, Syrian and Georgian, with Eritreans as the leading group. This is due to spouses who have been granted family reunification. Establishing a relationship with Eritrean refugees by exchanging knowledge on their rich textile history was more difficult than with other refugee groups because of the language barriers. Tigrinya and its textile vocabulary as well as Eritrean culture are relatively unknown in Denmark. There is also a lack of items in the museum's collections from this area as far as we can see. So the following questions seemed relevant : What is the clothing still in use from the area? How has it evolved over time? How could this knowledge help improve the refugees situation and integration within the THREAD project?

3. Case study: Eritrean textile terminology

3.1 Exploring Tigrinian textile lexicography

Tigrinya (TN) is, along with a few others, a national language in Eritrea. Because of family history and

education, most refugees speak TN. Occasionally they have a small understanding of Arabic but rarely any people from this area speak English.

TN is a Semitic language (spoken by around 15 million people) which is related to Arabic in several manners. It is seen from both epigraphy and phylogenetic studies [Kitchen et al. 2009] that it derives from the Ethio-Semitic group through the Geez language from which it has its alphabet. The Geez language derives from South Semitic about 2850 years ago, itself deriving from Central Semitic that contains Arabic about 5400 years ago. This relationship is most visible in the verbal system and in some fundamental part of the basic vocabulary, like the Swadesh list of 100 or 200 most basic words. TN also has a strong Kushic influence, but a second wave of influence came later on from Arabic due to continuous exchanges between the coasts of Yemen and Eritrea. This is made possible because of their location facing each other across the Red Sea. Many technological terms are found in the Ethiopic languages that were borrowed from Arabic through the ages [Leslau 1956]. Naturally, textiles are deeply concerned by this influence and therefore studying textile vocabulary can be challenging from a philological and etymological perspective [Bulakh Kogan 2017]. Also many dictionaries lack this information. A deeper knowledge of TN, Geez, and Arabic is necessary to track the words and origins. This vocabulary is related to the archeological and ethnographic issues.

3.2 Exploring tigrinian textile lexicography

The textile history might be analyzed through vocabulary studies and dialectal variations. We undertook a historical and linguistic review by a study of relevant literature, related to the Tigre region and costume/clothing traditions. What was especially important was Leslau's collection of texts [Leslau 1941], a Tigrinya-English dictionary [Kane 2000], and a Geez dictionary [Leslau 1987]. Each concept/term was searched for in these references as well as their etymology and roots questioned. Gervers's studies [Gervers 2003, 2017] on clothing was an important source to extract these concepts and terms in order to establish the categories of the ontology and diachrony presented above. We thank Prof. Alessandro Gori, who directed the IslHornAfr ERC project at the Saxo Institute at that time, for helping us identifying these sources.

In order to learn more about TN, we searched for members of the Eritrean community in Denmark and found refugee volunteers willing to participate in the project. This was located in the Netværkshuset in Gentofte, a communal association organizing activities for the refugees in Copenhagen. Christian Gaubert began lessons there – both written and oral - in Tigrinya. The lessons were given by Mikael Kahsay, a medical assistance student whom we warmly thank for the time and enthusiasm he spent in that task.

In addition to this learning experience, the information collection phase began. A 10 point questionnaire was set up that asked questions on family clothing knowledge in Danish and TN. This questionnaire was then widely

distributed in the community (centres, churches and restaurants). We planed to make follow-up interviews and look for non-verbal and visual representations of the concepts. Though the questionnaire campaign received little feedback and answers, we were given confirmation that many families keep traditional clothes, especially dresses, with them and that some of them were willing to share their knowledge.

3.3 A preliminary result: an overview of the influences on textile terminology and lexicography in Tigrinya.

We will develop here the example of the basic verbs used related to textile production and clothing. These cover the following verbs in English : to card, to spin, to weave, to embroider, to cloth, to wear.

- to card (cotton, wool) : Geez *bazzata* card/eviscerate > **TN** በዘተ *bäzzätä*.
- to spin : Hebrew *pāṭal* twist, Arabic *fatala* twist, spin > Geez *fatala* > **TN** ፈታሊ *fätälä* - *fätli* cord, fiber > Cushic *fatal* to spin.
- to weave : South Arabic *'nm* weavers > Geez *'anama* to weave > **TN** ኣነመ *'anämä* ኣለመ *alämä*, Amh. *Anämä*. *alämä* has also the meaning 'to plot', like in latin proverbs and many lain rooted langages. See also Aramean *šepar* braid > Arabic *ḡafara* > Geez *ḡafara* braid/weave/plot > **TN** ሻፋራ *šāfārä*, Amharic *taffārä*.
- to embroider : Geez *ṭalafa* snatch away, kidnap > **TN** ጠለፈ *ṭäläfä* to embroider (also to abduct, intercept, tap, hijack).
- to cloth : Semitic, Arabic. *kadana bi* put on clothes, protect > Geez *kadana* to cloth > **TN** ኣደኣ *kädänä* to cloth; *kadan* cover/cloth > Bilen *kadän* cover; ኣኣደድደድ *'akkädadäna* ways of clothing.
- to wear : Akkadian *labāšu*, South Arabic *lbs*, Arabic *labisa* wear > Geez (*'a*)*lbsa* wear > **TN** ለበሰ *läbäsä*.

These elements are summarized in a table which samples some basic influences that can be traced on TN since late antiquity through the successive reigns in northern and coastal Ethiopia, from the pre-axoumite period to the italian colonization. The vocabulary is divided into several categories : material, verbs, textile or textile-made objects.

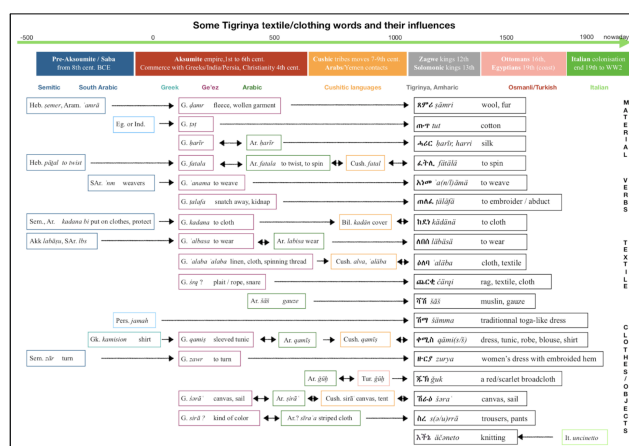


Fig. A diachronic diagram explaining textile words influence in Tigrinya

4. Conclusion and future perspectives

This project, that consists of the study and survey of textile vocabulary conducted with refugees, seemed to us to be beneficial on several levels. We felt that both the women who participated in the training, and our helpers at the Community Refugee Centre, were deeply enthusiastic about sharing elements of their knowledge in an academic context. They also could benefit from this new network of institutional and personal contacts. We also concluded that new research opportunities could emerge from this collaboration. However, they require a considerable investment of time and scientific expertise that can only be achieved through extensive collaboration.

As a next phase we could exhibit the costumes/clothing and map them geographically as well as explaining the techniques. We would also like to deepen the vocabulary collection with several initiatives including the exploration of written sources (literature, web) with refugees.

5. Bibliographical References

- Bulakh, Maria, Leonid Efimovich Kogan. 2017. The Arabic-Ethiopic Glossary by Al-Malik Al-Afdal: An Annotated Edition with a Linguistic Introduction and a Lexical Index. Handbook of Oriental Studies. Section 1 The Near and Middle East, Volume: 113. doi:10.1163/9789004321823
- Gervers, Michael. 2003. « Clothing ». Encyclopaedia Aethiopica ed. Sigbert Uhlig. Wiesbaden: Harrasowitz.
- . 2017. « Clothing ». In Ethiopia: History, Culture and Challenges, Berlin, 2017, p.25 and 149-53.
- Kane, Thomas Leiper. 2000. Tigrinya-English dictionary. Springfield, VA: Dunwoody Press.
- Kifle, Nazareth Amleson and Andersen, Øivin. 2008. Terminological challenges in the Norwegian-Tigrinja Dictionary in NORDTERM Proceedings Kunnskap og fagkommunikasjon, Bergen 13.-16.6 2007, p.117-130
- Kitchen, A., Ehret, C., Assefa, S. and Mulligan, C. J., 2009. Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East, Proc. R. Soc. B, doi:10.1098/rspb.2009.0408.
- Lervad, Susanne 2016. Professional Non-verbal Communication in the Field of Textiles in Représentation verbale et nonverbale en Terminologie, Journée d'Etude Toth, 2013, Copenhagen, Centre for Textile Research, p. 177-183.
- Leslau, Wolf. 1941. Documents tigrigna (éthiopien septentrional), grammaire et textes. Collection linguistique 48. Paris.
- . 1956. « Arabic Loanwords in Tigrinya ». Journal of the American Oriental Society 76 (4): 204–213. doi:10.2307/596147.
- . 1987. Comparative dictionary of Ge'ez (Classical Ethiopic), Ge'ez-English-English-Ge'ez with an index of the Semitic roots. Wiesbaden: Harrassowitz.

Svårt med facilitering? – om terminologiska arbetsgrupper, konsensus och terminologens roll

Henrik Nilsson

terminolog

C.A.G Next

Kungsgatan 37, SE-111 56 Stockholm, Sverige

0046729614682

henrik.nilsson@cag.se

I några länder är ”terminolog” ett erkänt yrke, i andra mer okänt. Terminologiarbete utförs dock av många, oftast i grupp och ofta med en medverkande terminolog. Det speciella för terminologen är att denna person oftast själv inte har expertis på det fackområde som behandlas utan i stället fungerar, med modern terminologi, som en facilitator. Men också som en ”utvinnare”, en medlare och en katalysator. Hur går detta till och vad är det egentligen terminologen gör under och mellan möten? Vilka är de vanligaste problemen, lösningarna – och missuppfattningarna om denna yrkesroll? Hur kan terminologen fungera som katalysator – och vad händer om terminologen tar på sig eller får en ”diktatorroll”?

Detta föredrag kommer att, utifrån praktiska projekt, presentera (själv)reflektioner kring terminologens yrkesroll, funktion och beteende men också undersöka, utifrån teorier om konsensuskapande, arbetsprocesser i grupper där olika personer med varierande expertis medverkar. Föredraget kommer också att beröra vad terminologens expertis består av, hur terminologrollen har förändrats i takt med att terminologiarbetet vidgats i riktning mot ontologi- och kunskapshantering, modellering och mappning, och kort diskutera hur detta (bör) avspeglas i utbildning av terminologer.

Lagringsvaror men inga lagervaror – om europeiska nationella termbanker och deras programvaror

Henrik Nilsson

ordförande

Europeiska terminologiföreningen (EAFT)/C.A.G Next

Kungsgatan 37, SE-111 56 Stockholm, Sverige

0046729614682

henrik.nilsson@cag.se

I samband med Nordterm 2019 ordnar Europeiska terminologiföreningen (EAFT) ett uppföljningsseminarium, TERMINTRA II, till det seminarium som hölls i Oslo 2012, om nationella termbanker. Sedan 2012 har fler länder skapat nationella termbanker eller utvecklat de befintliga, medan andra ännu inte har en enda samlande termbank på nationell nivå. Vad kan vi lära av varandra när det gäller just den här typen av termbank? Och åt vilket håll bör den utvecklas – mer eller mindre användarmedverkan, mer eller mindre normativt innehåll, enklare återbruk av innehållet?

De tekniska lösningar som valts för just nationella termbanker verkar variera stort, vilket i sig är intressant. I takt med att dagens terminologihanteringssystem (TMS:er) utvecklas borde de vara ett alternativ när stora termbanker skapas. Klara de inte de krav som just den nationella termbanken och dess vida målgrupp ställer? Eller handlar det helt enkelt om att det är enklare att få pengar för att skapa en egen programvara? Kan det handla om

konkurrens från andra, öppna lösningar av wiki-typ?

Dessa frågor kommer att diskuteras i detta föredrag som kommer att ta upp exempel på nationella termbanker i Europa och diskutera deras val av tekniska lösningar. Föredraget kommer också att beröra de uppdateringar som är på gång av relevanta ISO-standarder som rör termbanker och termbanksutveckling.

Referenser

Nilsson, H. & Cuadrado, S. 2015. Software – Hard for National Termbanks? IITF Workshop. I: Lušicky, V. & Budin, G. (red.), *Book of Abstracts of the 20th European Symposium on Languages for Special Purposes, 8–10 July 2015*, Vienna, Austria, Multilingualism in Specialized Communication: Challenges and Opportunities in the Digital Age, ISBN: 978-3-200-04186-8, <https://fedora.phaidra.univie.ac.at/fedora/get/o:406394/bdef:Content/download>, 2019-02-15

Hur mycket språkteknologi behövs det i terminologiutbildningen?

Niina Nissilä & Anita Nuopponen

Enheten för marknadsföring och kommunikation

Vasa universitet

niina.nissila@univaasa.fi, anita.nuopponen@univaasa.fi

Nyckelord: terminologiundervisning, terminologiförvaltning, språkteknologi

I vårt bidrag diskuterar vi förnyande av terminologiutbildningen vid Vasa universitet inom läroämnet kommunikationsvetenskaperna. Under de fyrtio åren som vi har haft terminologikurser har innehållet för det mesta fokuserat i de grundläggande terminologiska och begreppsanalytiska metoderna och principerna. På kandidatnivån (VIEK) riktar sig en grundkurs i terminologi till alla kommunikationsstuderande, på magisternivån har vi en kurs i begreppsanalys både i det multidisciplinära magisterprogrammet för kommunikation (VIMA) och i magisterprogrammet för teknisk kommunikation (TEVI). Därutöver har vi en specialisering i terminologi i det sistnämnda.

I detta föredrag fokuserar vi speciellt i utvecklingen av en kurs i terminologiarbete och -förvaltning som

kommer i fortsättningen att fungera som en fortsättnings- och specialiseringskurs på båda nivåerna (VIEK & TEVI). Målet är att studenterna ska känna igen terminologiska problem och kunna argumentera för behovet för terminologiarbete och -förvaltning i arbetslivet, och veta hur terminologins verktyg kan användas. På kursen behandlas även kriterierna för termbildning och -evaluering. Vidare ska studenterna bekanta sig med de grundläggande principerna för språkteknologiska applikationer som kan användas i terminologihantering och kunna utvärdera dem. Eftersom fokus i deras utbildning ligger i teknisk och fackkommunikation i stället för språk och översättning, letar vi efter ett svar på frågan vilka är de mest relevanta språkteknologiska element som ska ingå i kursen.

Terminologi på tværs af Danmark – og EU

Om eTranslation TermBank - indsamling af terminologi til maskinoversættelse

Olsen, S., Henriksen L.

Center for Sprogteknologi, NorS, Københavns Universitet
{saolsen, linah}@hum.ku.dk

Abstract

Denne artikel handler om eTranslation TermBank-projektet med fokus på arbejdet udført af den danske partner. Først beskrives formålet med projektet, nemlig indsamling af terminologi til EU's maskinoversættelsessystem og motivationen for denne indsamling. Dernæst rapporteres om terminologiforvaltningen i de offentlige institutioner i Danmark. Herefter gennemgås resultaterne af indsamlingen af både dansk, norsk og ungarsk terminologi, og til sidst beskrives en ansøgning om en videreførelse af projektet.

Keywords: keyword1, keyword2, keyword3

1. eTranslation TermBank-projektet

eTranslation TermBank (eTTB) var et EU-projekt der løb fra september 2017 til februar 2019. Projektets formål var at indsamle terminologiske ressourcer med fokus på nogle specifikke fagområder for alle officielle EU-sprog samt for norsk og islandsk. Der skulle kun indsamles terminologi der ikke i forvejen var tilgængelig online, og som havde en licens tilknyttet (Gornostaja et al., 2018).

eTTB-projektet udsprang af European Language Resource Coordination (ELRC)¹, et netværk og en dataportal der har til formål at øge opmærksomheden omkring sproglige ressourcer hos offentlige institutioner i EU's medlemslande samt indsamle og klargøre sproglige ressourcer til EU's oversættelsessystem eTranslation (Lösch et al., 2018)². ELRC fokuserer især på at indsamle parallelle ressourcer, terminologi inkluderet; det sidste dog med begrænset succes, hvilket blev motivationen for eTTB.

1.1 Fagområder og online-tjenester

De fagområder som indsamlingen fokuserede på, er tæt knyttet til tre online EU-tjenester (DSIs - Digital Service Infrastructures)

- Sundhedsportal – eHealth
- Retsinformation – e-Justice
- Onlinetvistbilæggelse – Open Dispute Resolution

Disse online-tjenester er igangsat under programmet Connecting Europe Facility (CEF), et forskningsprogram der investerer i og støtter opbygningen af digitale infrastrukturer. Onlinetjenesterne kan udnytte maskinoversættelse, og eTranslation, som er en videreudbygning af EU's tidligere maskinoversættelses-system, skal beriges med terminologi fra de tre ovennævnte fagområder for at sikre en bedre oversættelseskvalitet.

1.2 Sprog og partnere

Projektet havde 9 partnere fordelt på 8 lande og blev ledet af oversættelses- og lokaliseringsvirksomheden Tilde fra Letland som har stor erfaring med at lede EU-projekter.

Den danske partner var Center for Sprogteknologi, NorS på Københavns Universitet.

Da der skulle indsamles terminologi for alle EU-sprog samt for norsk (bokmål og nynorsk) og islandsk, var hver partner ansvarlig for at indsamle terminologi for mere end ét sprog. Den danske partner skulle indsamle terminologi for både dansk og ungarsk og skulle sammen med den islandske partner indsamle ressourcer for norsk.

Fokus var på at indsamle bilingval terminologi, men monolingval kunne også anvendes.

2. Maskinoversættelse og terminologi

Baggrunden og motivationen for at indsamle terminologi til maskinoversættelse er at maskinoversættelse ikke altid er i stand til at oversætte fagsprog og terminologi korrekt.

Nedenstående eksempler er venligst lånt af professor Bodil Nistrup Madsen og viser med al tydelighed hvor galt det kan gå når medicinske termer bliver oversat automatisk.

Dansk (fra begrebsbasen http://sundhedsdata.item.dk)	Definition	Engelsk (Google)
engangsdosering	"enkeltstående dosering af én lægemiddeldosis "	prefilled <i>Alternativ:</i> once a day regimen
pausering	"midlertidig afbrydelse af lægemiddel- ordination efter anvisning af en dertil autoriseret person"	break ring

Tabel 1. Eksempler på automatisk oversættelse af danske sundhedstermer.

¹ <http://lr-coordination.eu/>

²

<https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eTranslation>

Når en fagsproglig tekst skal maskinoversættes, kan man vælge at lade fagområdets terminologi få forrang over almensproglige ord og andre fagområders terminologi. Fx kræver oversættelsen af 'blank cartridges' at udtrykket findes i systemet som en bilingval term da det ellers vil blive oversat til 'blanke patroner i stedet for 'løst krudt'. Den monolingvale terminologi bruges, i forbindelse med maskinoversættelse, hovedsagelig til præprocessering, domænebestemmelse og sprogmodellering.

3. Terminologi i Danmark

I modsætning til mange af vores nabolande, findes der i Danmark ingen central organisation der forvalter terminologi.

3.1 Terminologiske initiativer og netværk

Der har gennem årene været flere større terminologiske initiativer, hvoraf det største nok var DanTermCentret (1998-2015) der arbejdede med at informere om og oplære sproglige medarbejdere i systematisk behandling af terminologi. DanTermCentret lukkede i 2015.

På Copenhagen Business School var der i mange år forskning i terminologi på højt plan – samt uddannelse af terminologer. Afdelingen er nu lukket.

Undervisning i terminologi i Danmark foregår – så vidt vi er orienterede – nu kun på Syddansk Universitet i Kolding.

Center for Sprogteknologi på Københavns Universitet har tidligere deltaget i andre EU-terminologiprojekter, hvor bl.a. termbankerne IATE og EuroTermBank blev udviklet.

Skønt der ikke findes en central organisation for terminologi, og skønt forskningen inden for feltet er blevet decimeret, findes der dog et par landsdækkende terminologinetværk i Danmark: FORVIR og Terminologi-gruppen.

Forum for Videnmodellering i Offentligt Regi – FORVIR, er et netværk for ansatte i den offentlige administration. Formålet med netværket er bl.a. at styrke det terminologiske begrebsarbejde både kvantitativt og kvalitativt.

Terminologigruppen er et netværk for både den private og offentlige sektor. Formålet er bl.a. at udveksle erfaringer og deltage i nordisk og internationalt terminologiarbejde.

Sluttelig er det vigtigt at understrege at den danske regering i 2018 nedsatte Det Sprogteknologiske Udvalg der bl.a. havde til formål at afklare behovet og perspektiverne for en national termbank, Kirchmeier et al., 2019.

3.2 Terminologiarbejde i danske offentlige institutioner

Nogle offentlige institutioner i Danmark udfører grundigt terminologi- og begrebsarbejde.

Dette gælder fx Sundhedsdatastyrelsen som både arbejder med klassifikationer til brug for det danske sundhedsvæsen og med sundhedsterminologi der kan bruges som redskab for entydig kommunikation på tværs af faggrupper i sundhedsvæsenet.

Også Socialstyrelsen arbejder med terminologi og har en offentlig tilgængelig database: socialebegreber.dk samt en

række foldere og et e-læringsmodul der giver en introduktion til begrebsarbejdet.

Derudover har Domstolsstyrelsen udarbejdet en ordbog, en samling af juridiske begreber med let forståelige definitioner og forklaringer.

Der findes andre offentlige og halvoftentlige institutioner som arbejder med terminologi og begreber, fx har LO udgivet en arbejdsmarkedsordbog over fagforeningsrelaterede termer. Det er dog langt fra normen i danske institutioner at udføre begrebs- og terminologiarbejde. Kun 30 % af de institutioner vi henvendte os til, arbejdede med terminologi i et eller andet omfang. Ca. 70 % rapporterede at de ikke beskæftigede sig med terminologi overhovedet, men at større oversættelsesopgaver blev sendt til oversættelsesbureauer, uden krav om at få terminologien tilbage fra bureauet, og ellers brugte de Google. Dette understøttes af rapporten fra Det Sprogteknologiske Udvalg, hvor det nævnes at 80-90 % de danske offentlige institutioner udliciterer deres oversættelsesarbejde til oversættelsesbureauer.

4. Resultat af indsamlingen

Projektet identificerede i alt 1178 terminologiske ressourcer og endte med at levere 241 ressourcer, processeret og klargjort til maskinoversættelse, mens 131 ressourcer blev leveret uden at blive processeret i projektets løbetid. De øvrige identificerede ressourcer blev enten ikke indsamlet af forskellige grunde, eller deres licensforhold var stadig uafklarede da projektet udløb.

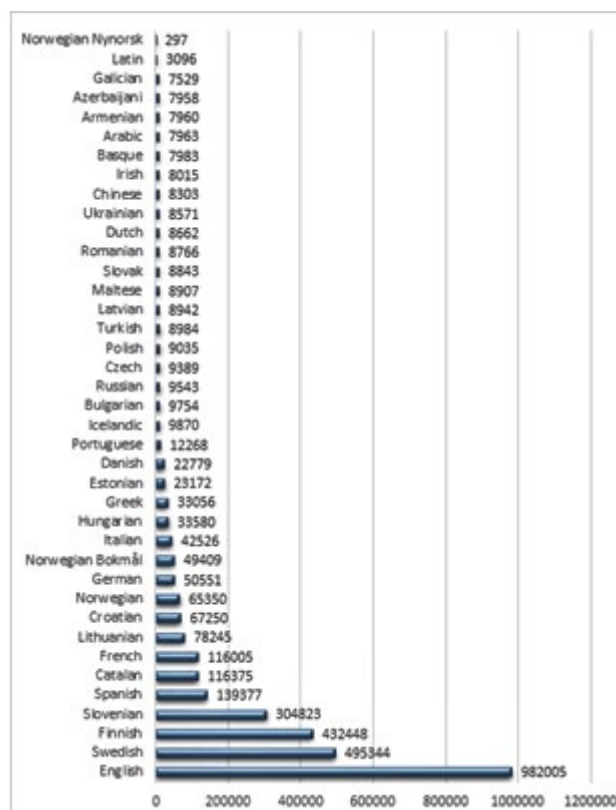


Fig.1 Figuren viser antallet af indsamlede og processerede termer for hvert sprog.

På figur 1 kan man se fordelingen af antal ressourcer pr sprog, bemærk at der også er blevet indsamlet ressourcer for ikke-EU-sprog, såsom arabisk og russisk. Dette skyldes som oftest at det pågældende sprog er en del af en multilingval ressource.

På figur 2 kan man se de forskellige fagområder de indsamlede ressourcer dækker. Som det fremgår, dækker de indsamlede ressourcer et større antal fagområder end blot de tre som projektet fokuserede på.

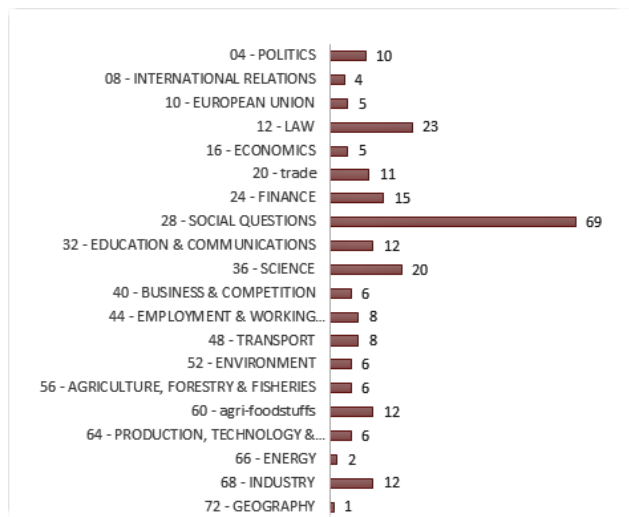


Fig. 2 Antallet af processerede ressourcer fordelt på hoved-fagområder.

4.1 Indsamling af terminologi i Danmark

I Danmark henvendte vi os til ca. 40 institutioner og har fået ressourcer fra 8 af disse.

Vi indsamlede 23.000 danske termer fra bilingvale ressourcer, og ca. 20.000 termer fra monolingvale ressourcer. Termerne er fordelt på 9 samlinger og var fordelt således på fagområder:

- Medicin: 35.000 termer
- Onlinetvistbilæggelse (ODR): ca. 8.000 termer

4.2 Indsamling af terminologi i Norge

Ud over Danmark har også den islandske og den svenske partner i projektet deltaget i indsamlingen af norsk terminologi.

I alt er der indsamlet 115.000 norske termer (bokmål og nynorsk til sammen) fra bilingvale ressourcer samt et ukendt antal monolingvale ressourcer. Termerne er fordelt på 10 samlinger. Vi kender ikke de præcise tal for de forskellige fagområder.

4.3 Indsamling af terminologi i Ungarn

Det var som udgangspunkt meget vanskeligt at identificere ungarske terminologiske ressourcer, men med hjælp fra mange partnere i projektet, endte vi med at identificere 10-12 institutioner som er indehavere af

terminologieressourcer. Det samlede udbytte blev 33.000 ungarske termer, flest bilingvale og flest inden for fagområdet onlinetvistbilæggelse.

5. Konklusion og fremtidige initiativer

Vi må konkludere at det er et tidskrævende arbejde at indsamle terminologi – selv fra offentlige institutioner der if. PSI-direktivet³ har pligt til at udlevere data som ikke er personfølsomme eller på anden måde særligt beskyttede.

Der er i Danmark i offentlige institutioner generelt meget lidt fokus på sproglige data og terminologi, og hvad de kan bruges til. Langt de fleste beskæftiger sig som sagt slet ikke med sprog, og de der gør, havde ofte mange betænkeligheder ved at skulle udlevere deres termressourcer.

De offentlige institutioner stillede mange spørgsmål og udviste megen usikkerhed om anvendelsen af data. De hyppigste spørgsmål var

- Hvem vil bruge terminologien og hvordan?
- Hvor lagres den?
- Kan data blive misbrugt?
- Kan vores data ende med at gøre mere skade end gavn?
- Når vores data opdateres hos os, hvordan sikrer vi at data også opdateres i EU-systemet?

De største udfordringer var dels det manglende kendskab til eTranslation, dels manglende viden om PSI-direktivet. Hvis man ikke kender til maskinoversættelse og eTranslation, er interessen for at forbedre systemet med egne data naturligvis mindre.

For at imødegå disse udfordringer har holdet bag eTTB-projektet udarbejdet en ny ansøgning inden for CEF-programmet. Denne ansøgning fokuserer dels på en terminologiportal med opdaterede termer, dels på at informere om eTranslation-systemet.

Terminologiporten skal opbygges som et 'federated network', dvs. bestå af links mellem institutionernes termdata og portalen. Dette sikrer at en opdatering af data i hjeminstitutionen også vil slå igennem i eTranslation TermBank-portal. Derved elimineres et af institutionernes største forbehold over for eTranslation TermBank.

Projektet vil desuden via workshops o.lign. øge kendskabet til eTranslation hos de offentlige institutioner i EU, og hvordan de kan anvende og have gavn af systemet, og samtidig udbrede viden om vigtigheden af terminologi i maskinoversættelse samt informere om den nye termportal.

6. Referencer

Kirchmeier, S., Henrichsen, P. J., Diderichsen, P. og Hansen, N., B., 2019. *Dansk Sprogteknologi i Verdensklasse*. Dansk Sprognævn.
<https://dsn.dk/udgivelser/sprognaevnets-udgivelser/sprognaevnets-rapporter/sprogteknologi-i-verdensklasse>

Lösch, A., Mapelli, V., Piperidis, S., Vasiljevs, A., Smal, L., Declerck, T., Schnur, E., Choukri, K. og Van

³Lov om videreanvendelse af den offentlige sektors informationer :

<https://www.retsinformation.dk/forms/r0710.aspx?id=29235>

Genabith, J. (2018). European Language Resource Coordination: Collecting Language Resources for Public Sector Multilingual Information Management. In Nicoletta Calzolari et al., editors, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, p.1339-1343, Miyazaki, Japan. European Language Resources Association (ELRA).

Gornostaja, T., Auksoriūtė, A., Dahlberg, S., Domeij, R., van Dorrestein, M., Hallberg, K., Henriksen, L., Kallas, J., Krek, S., Lagzdīņš, A., Lilles, K., Mitkevičienė, A., Olsen, S., Pedersen, B. S., Pesliakaitė, E., Povlsen, C., Repar, A., Rozis, R., Sauberer, G., Thorbergsdóttir, Á., Vasiljevs, A., Vasiļevskis, A., Vaus, M., Zabarskaitė, J. 2018. eTranslation TermBank: stimulating the collection of terminological resources for automated translation. In *Proceedings of the XVIII EURALEX International Congress, Book of abstracts*, Ljubljana, Slovenien.

Digitalisering av offentliga tjänster – begreppsarbete som en del av utveckling och digitaliseringen av statsunderstödsverksamhet

Katri Seppälä, Mari Suhonen

Terminologicentralen TSK, Runebergsgatan 4c B 20, FI-00100 Helsingfors
katri.seppala@tsk.fi, mari.suhonen@tsk.fi

Abstrakt

I Finland har finansministeriet som resultat av en förutredning inlett ett omfattande projekt för utveckling och digitalisering av statsunderstödsverksamheten. Projektet ska främja samarbetet mellan de statliga myndigheter som beviljar understöden. Digitaliseringen förverkligas genom att skapa en enhetlig och kundinriktad nationell modell för datasystem och tjänster för myndigheternas gemensamma bruk. Målet är att minska mängden administrativt arbete i bidragsprocessen och den administrativa bördan även för dem som ansöker om understöd.

Begreppsarbetet är en viktig del av utvecklingsprojektet, eftersom en gemensam verksamhetsmodell och ett enhetligt ärendehanteringssystem förutsätter att alla parter har en gemensam uppfattning om de centrala begreppen. De existerande begreppen måste redas ut innan själva utvecklingen påbörjas så att alla vet vilka begrepp diskussionen rör. Men begreppsarbetet behövs också vid själva utvecklingen, för det kommer att uppstå nya begrepp. Eftersom projektet har många parter och även koppling till andra utvecklingsprojekt, krävs det inte bara expertis på begreppsanalys utan också förmåga att koordinera det omfattande samarbetet så att det är möjligt att nå högklassiga resultat.

Nyckelord: begreppsarbete, digitalisering, statsunderstödsverksamhet

1. Utvecklingsprojektet för statsunderstödsverksamheten

Digitalisering av statsunderstödsverksamheten är ett omfattande projekt eftersom nästan alla landets ministerier är delaktiga i verksamheten. Det gäller alltså likaväl statsunderstöd för flygfälten, brandsläckningsutrustningar, rekreationsområden, stödbostäder för specialgrupper, museer, konstnärer, fritidsaktiviteter för barn, utvecklingssamarbete och så vidare. Även om statsbidragsmyndigheterna följer statsunderstödslagen har deras verksamhet särdrag. Det är en grund för att det har under åren uppstått olika tolkningar också om de centrala begreppen som borde vara gemensamma. Dessutom föreskrivs vissa myndigheternas verksamhet närmare i andra författningar där det eventuellt används olika termer för samma begrepp som är i bruk i statsunderstödslagen. Därför är den gemensamma statsunderstödslagen inte en fullständig källa för begreppen och termer utan man måste reda ut hur de olika myndigheterna brukar använda begreppen, hitta kärnan i begreppet som alla parter accepterar och harmonisera begreppen så mycket som möjligt.

Under 2017 gjorde undervisnings- och kulturministeriet en förhandsutredning för digitaliseringen av statsunderstödsverksamheten. Den innehöll en beskrivning av situationen inom verksamheten samt vision för utveckling av verksamheten med förslag till åtgärder. I detta sammanhang dök den ovannämnda utmaningen med de centrala begreppen upp. Därför bildade undervisnings- och kulturministeriet med stöd från CSC (Centret för IT-expertis) en arbetsgrupp med medlemmar från olika statliga myndigheter och startade begreppsarbetet med Terminologicentralen då förberedningen av utvecklingsprojektet fortsatte våren 2018.

I 2019 blev förberedningsarbetet färdigt, finansministeriet tog över och inledde i början av mars ett femårigt utvecklingsprojekt för reform och digitalisering av statsunderstödsverksamheten. Begreppsarbetet fortsätter under det första året av utvecklingsprojektet.

Det viktigaste målet för projektet är att skapa en gemensam verksamhetsmodell och ett enhetligt datasystem. Den nya digitala kundorienterade processen ska omfatta hela statsförvaltningen och täcka alla faser som ingår i statsunderstödsverksamheten: riktande, sökning, beviljande, uppföljande och utvärdering. Den ska bygga på den gemensamma kärnan, men ta också hänsyn till särdrag i olika myndigheternas verksamhet.

När ett gemensamt ärendehanteringssystem är tillgängligt, kan myndigheterna spara resurser, eftersom alla behöver inte investera i byggande och uppdatering av egna system. Det gemensamma systemet främjar också det praktiska samarbetet mellan de statliga myndigheter som beviljar understöden och myndigheterna får stöd från varandra. För dem som ansöker understöd är ansökningsprocessen lättare, eftersom de behöver inte lära sig att använda flera system, om de ansöker stöd för flera ändamål eller från flera myndigheter. Både myndigheter och deras kunder kan alltså arbeta mer effektivt.

2. Begreppsarbetets roll i utvecklingsprojektet

Den harmoniserade begreppsinformation som är resultatet av begreppsarbetet kan utnyttjas både i planering och genomföring av utvecklingsprojektet. För att bygga basen för samarbetet, behövs enighet om de gemensamma begreppen. Om en term tolkas på flera sätt eller samma term används för att hänvisa till olika

begreppsinnehåll, finns det ingen fungerande bas för utvecklingsarbetet.

Harmonisering kräver diskussion och tar sin tid. Då begreppsarbetet är en del av en utvecklingsprojekt, är det alltså viktigt att inleda begreppsarbetet redan i förberedningsfasen. På det sättet är det möjligt att få tillräckligt tid för arbetet innan själva utvecklingen börjar och att undvika bortkasta det arbete som luddiga begrepp skulle förorsaka.

När begreppen är utredda är det lättare att bygga ett datasystem som är logiskt och användarvänligt. Om det behövs överföra information från ett system till ett annat, kan man med hjälp av definitioner försäkra semantiskt interoperabilitet mellan systemen och minska risken för misstolkning av information. Samma gäller för kommunikation mellan människor.

Ett logiskt och enhetligt datasystem som utnyttjar harmoniserade begrepp förbättrar rättskyddet hos dem som ansöker, eftersom det är lättare att använda. Myndigheterna i sin tur sparar arbetstid när de som ansöker behöver mindre hjälp med ansökningen. Myndigheter kan också samla enhetlig information med hjälp av systemet. I fortsättningen hjälper analysering av samlad information myndigheter att mer effektivt följa upp och utvärdera verksamheten, vilket ökar verksamhetens genomslagskraft och transparens.

3. Begreppsarbete i praktiken

Terminologiprojektet har utförts i tre skeden. Från och med början har vi haft en arbetsgrupp med jurister och andra tjänstemän från olika ministerier och ämbetsverk. Terminologicalcentralen önskade 5–7 medlemmar i arbetsgruppen men flera organisationer ville delta och därför har vi tagit med alla experter som vill vara med. Det har varit sammanlagt cirka 20 experter som har deltagit. I det senaste skedet har vi också haft medlemmar från centralorganisationer för de som söker understöd eftersom dessa begrepp rör dem också.

Den stora gruppen fungerar bra eftersom terminologen behöver inte vara en ordförande utan vi har medlemmar från CSC (Centret för IT-expertis) och undervisnings- och kulturministeriet som tar hand om att leda diskussion och ta hand om tidtabellen i mötet. Gruppen har möter varannan vecka, tre timmar per möte. Dessutom har vi haft möten med deltagarna från CSC och kultur- och undervisningsministeriet där vi förbereder materialet.

4. Att välja begrepp

En stor fråga i begreppsarbete är hur man väljer begreppen. I Finland har vi statsunderstödslagen som är den främsta källan till flera begrepp. Dessutom finns vissa andra lagar och förordningar som gäller

statsunderstöd och man följer också förvaltningslag vars syfte är att genomföra och främja god förvaltning allmänt.

Statsunderstödslagen är en bra källa till vissa begrepp som är gemensamma till alla som beviljar statsunderstöd. Till exempel vissa slag av statsunderstöd, liksom *yleisavustus* (allmänt stöd) och *valtionavustus* (statsunderstödsbeslut). Statsunderstödslagen är dock på vissa delar otillräcklig eller kan tolkas på flera sätt. Lagen trädde i kraft i 2001 och den har tillämpats på varierande sätt under åren. Till exempel när vi har i begreppsarbetet harmoniserat de olika slag av statsunderstöd hos de olika organisationerna har vi gjort en annan typ av hierarki än i den klassificering som framställs i lagen. Orsaken är oftast att begreppen som finns i lagen är inte begränsade på ett entydigt sätt. Ett annat exempel är bedömning av verkningarna av statsunderstöd som beskrivs mycket kort i lagen och som har flera tolkningar i Finland och även flera jämförelsepunkter internationellt.

När man vill beskriva den statsunderstödsverksamhet som genomförs i praktiken räcker dock inte lagen som en källa. Lagen ger till exempel inte namn på alla handlingar, aktörer och verksamhet som man behöver i praktiken även om dessa baserar sig på lagen. Därför har olika organisationer eller avdelningar inom en organisation använt olika termer och då behövs samarbetet mest. Arbetsgruppen måste kunna identifiera begreppen och rekommendera termer som kan användas i fortsättningen åtminstone i den gemensamma kommunikationen.

En mindre men märkvärdig detalj är att vissa understöd som beviljas av *Europeiska unionens fonder för inrikes frågor* måste använda de termer som kommer från EU:s förordningar. Där finns inte termen *valtionavustus* utan stödet kallas för *valtioneutuk*. Det leder till en svår situation eftersom *valtioneutuk* i Finland betyder annat än *valtionavustus* och borde inte användas för detta begrepp. Därför diskuterar vi om termen kunde harmoniseras till *valtionavustus* åtminstone i finska förordningar som gäller dessa EU-fonder.

Till sist har vi begrepp som inte beskrivs i författningar. Det är fråga om hur myndigheterna planerar beviljande av statsunderstöd och hur både myndigheten och mottagaren av stödet bedömer och utvärderar mottagarens verksamhet. I dessa fall vill man både harmonisera begrepp som till exempel de som utvecklar verksamhet eller forskare har skapat. Innehållet av dessa begrepp brukar variera enligt den som har skapat till exempel en bedömningsmetod och därför hittar man inte allmänna definitioner någonstans. Dessa begrepp har varit svåra eftersom de verkar allmänspråkliga men har ändå en specifik betydelse. Man brukar till exempel använda ordet *tulos* (resultat) i flera olika betydelser. Därför har vi rekommenderat mer exakta termer *välitön tulos* och *lopputulos* till två olika typer av resultat.

5. Semantisk interoperabilitet

I Finland har man under flera år arbetat för semantisk interoperabilitet inom den offentliga förvaltningen. Terminologiska ordlistor är en väsentlig källa för den semantiska interoperabiliteten. Den offentliga förvaltningens gemensamma begreppsarbete påbörjades redan under 2010. Flera aktörer har dock redan tidigare skapat terminologiska ordlistor på sitt område och fortsätter med det.

Befolkningsregistercentralen har ett verktyg¹ där man ska publicera och upprätthålla en ordlista över de gemensamma begreppen till den offentliga förvaltningen. Där ska också ämnesspecifika begrepp publiceras. Dessa begrepp ska kunna användas i till exempel datamodeller.

Terminologicentralen har sin TEPA-termbank² som innehåller terminologiska ordlistor och annat termmaterial från flera olika områden. Vi publicerar där till exempel ordlistor från de flesta av våra terminologiprojekt. TEPA är en bra källa om man vill undersöka gemensamma begrepp.

Författningarna är inte ordlistor men den elektroniska författningssamlingen³ är också en källa för semantisk interoperabilitet. Lagarna styr ju myndigheternas verksamhet och måste därför användas när man väljer eller skapar begrepp.

6. Exempel: handlingar och ärenden

Begreppssystemet i bilden 1. visar hur vi har undersökt och harmoniserat begreppen inom statsunderstödsverksamhet med begreppen inom den övriga offentliga förvaltningen. Begreppsarbetet görs tillsvärdare endast på finska och de svenska ekvivalenterna har valts bara för denna artikel.

Begreppet *handling* har definierats i vår ordlista på samma sätt som i den gemensamma ordlistan för den offentliga förvaltningen (i Sanastot-verktyget). Det ska användas i samma bemärkelse.

Ansökan, statsunderstödsbeslut, avtal om statsunderstöd, utredning och ärende som gäller statsunderstöd har för sin del gemensamma kännetecken med andra myndigheternas motsvarande handlingar men dessutom har de kännetecken som är specifika inom statsunderstödsverksamheten.

Vissa handlingar och ärenden är specifika för endast statsunderstödsverksamhet, som till exempel *ansökan om statsunderstöd* och *ärende som gäller beviljande av statsunderstöd*.

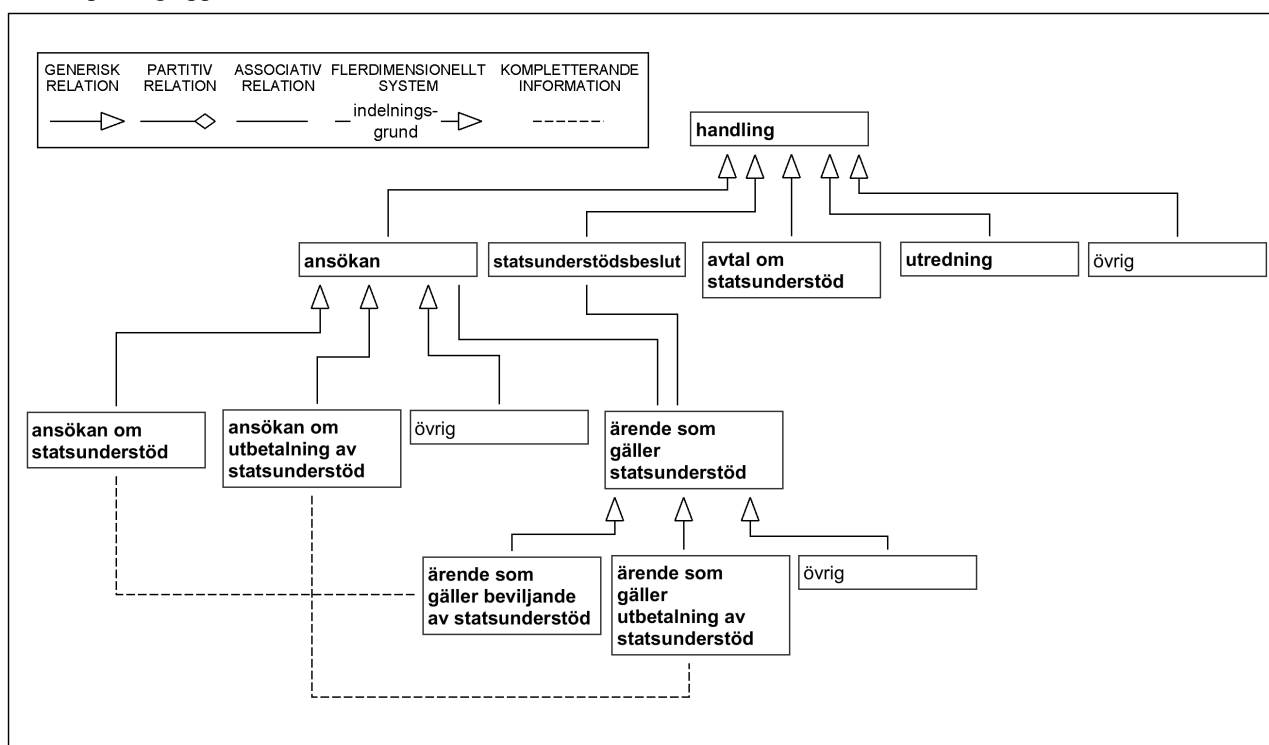


Bild 1: Handlingar

¹ Sanastot-verktyget <https://sanastot.suomi.fi/>

² TEPA-termbank <http://www.tsk.fi/tepa/sv/>

³ Finlex <https://www.finlex.fi/sv/>

7. Utveckling och utmaningar

Eftersom statsunderstödsverksamhet gäller hela statsförvaltningen och även EU:s författningar påverkar vissa stöd, diskuterar arbetsgruppen ofta vad som är gemensamt till alla eller generiskt på ett sätt som ska tas hänsyn till i begreppsarbetet.

Till exempel utrikesministeriet beviljar stöd som ska användas utanför Finland och dessa stöd kan vara mycket omfattande. Däremot beviljar till exempel Centret för konstfrämjande även mindre stöd till enskilda konstnärer. Arbetsgruppsmedlemmar måste därför vara färdiga att kompromissa så att alla för helhetens skull viktiga begrepp finns på ordlistan även om alla begreppen inte brukar användas i varje organisation.

En annan utmaning är de begrepp som är gemensamma men som har ingen tillsvidare har definierat. Det kan kännas frustrerande när man tror att det finns – eller att det borde finnas – någonstans en allmän definition till begreppet och därför vill man inte definiera begreppet i denna arbetsgrupp även om begreppet behövs för begreppssystemet. Detta ska bli lättare under tiden om den offentliga förvaltningens gemensamma ordlista utvecklas systematiskt.

En tredje utmaning gäller framtiden. Om man vill att vårt begreppsarbete ska utnyttjas efter projektet ska någon ta ansvar för att upprätthålla och utveckla ordlistan. Man ska också se till att ordlistan används aktivt. Sanastot-verktyget och modellen för förvaltning av de gemensamma begreppen ska vara viktiga.

Det roliga med arbetsgrupper är att experter vanligen är glada när de får diskutera sina begrepp och jämföra erfarenheter med kolleger. Experterna bearbetar inte bara en ordlista utan också lär från varandra och får kamrattstöd. I denna arbetsgrupp har experterna varit speciellt beredda att förstå varandra och vi har haft en mycket behaglig stämning.

8. Litteratur

Valtionavustusprosessin digitalisointia koskeva esiselvityshanke (2017). Valtionavustusten digitalisoinnin esiselvitys. Opetus- ja kulttuuriministeriön julkaisuja 2017:46. ISBN (PDF) 978-952-263-525-9.

Undervisnings- och kulturministeriet (2019). Utveckling och digitalisering av statsunderstödsverksamheten. <https://minedu.fi/sv/digitalisering-av-statsunderstoden>

Finansministeriet. Statsunderstödslag (2001). <https://www.finlex.fi/sv/laki/ajantasa/2001/20010688>

9. Språkliga resurser

Julkisen hallinnon ydinsanastoryhmä (2018). Julkisen hallinnon yhteinen sanasto. <http://uri.suomi.fi/terminology/jhs/>

Terminologicalentralen TSK (2019). TEPA-termbank. <http://www.tsk.fi/tepa/sv/>

When Automation Requires a Semantic Interface: Some Examples

Gerd Sjögren, Mats Granström, Henrik Andersen

Interverbum Technology, Copenhagen Translation

Stockholm/Sweden, Copenhagen/Denmark

{gerd.sjogren, mats.granstroem}@interverbumtech.com, henrik.andersen@copenhagentranslation.dk

Abstract

There are ambitious initiatives in sectors as diverse as Health, eCommerce and Process Industry to automate functions where semantic comprehension is a prerequisite for implementation. Without this comprehension, attempts to automate run the risk of becoming excessively time-consuming. The current presentation will focus on such cases. The main example is the Nordic Innovation-supported development project *Patient-Professional Communication Plugin (PPCP)*, where *free-text* extracts from an Electronic Health Record (EHR) can be displayed, and difficult terms «translated» into everyday language, in real time. The interface allows the user to choose these «translations», originating from a termbase, in one or several of the five major Nordic languages as well as English, Arabic, Somali and Tigrinya. Another example, using the same termbase, shows how someone writing a medical text can be prompted to use either a standardised medical term, or its everyday equivalent, rather than «professional jargon».

The examples have implications for how easy it is for a novice to tackle «controlled language», typically in the form of a drop-down menu, in situations where the individual does not recognise or fully understand the available alternatives. The presented examples make use of a terminology tool working in the background.

Keywords: Electronic Health Record, automation, semantic interface, standard medical language, terminology, language tool, e-commerce, controlled language

1. Background – health management objectives

There is a strong push in all Nordic countries, the EU and elsewhere for patients to become engaged in the management of their own health by giving them access to their own EHRs through sites such as Kanta/FI, Journalen/SE and Journal fra sygehus/DA. The health benefits of this engagement are well recognised, but so are certain problems. Health records are traditionally written by professionals for themselves and other professionals, and the language used is – or has typically been – a mixture of professional terms, jargon, intra-clinic abbreviations and similar. Understanding what the EHR says is not even limited to the patient and his/her immediate caretaker; it is often equally difficult for any other non-specialist health care provider.

Examples of communication challenges:

- Will patients understand their medical record online?
- Do health information sites understand the patient?
- Do Watson and other decision support systems speak any language beyond English?
- Is there a way to help health care professional to use standard medical language without effort?

To solve these challenges, there is a need to link professional expressions with everyday language. Bridging the gap between professional and non-professional terminology is in fact essential in ensuring that e-health applications are user-oriented and contribute to a high quality of care. Helping patients better understand their medical records involves expanding

acronyms and abbreviations, transforming idiolectic structures to standardised terms, and providing colloquial explanations without loss or distortion of information.

2. The project: developing and evaluating a pilot solution

The *Patient-Professional Communication Plugin (PPCP)*¹ project has created a technical infrastructure linking multilingual professional and non-professional language use. The platform is built through the integration of existing commercial tools for multilingual terminology management and text management. To provide a proof-of-concept for the technical solution, a pilot database specifically related to 'heart attack' has been developed. Heart attack is a potentially life-threatening diagnosis which is wide-spread and requires rapid and appropriate medical attention. Research shows that large patient categories may be underserved due to various reasons where communication is likely to play an important role.

The PPCP technical platform is intended for written and spoken digital exchanges. The fundamental approach of PPCP is to **ensure accuracy and validity** in the translation between professional and non-professional language. Until recently, machine translation systems such as Google Translate have primarily relied on statistical

¹ A Nordic Innovation-supported project formally named *Bridging the language gap : Creating a platform for patient-professional communication*, see <https://www.nordicinnovation.org/programs/patient-professional-communication-plugin-ppcp>

frequency, which – as we know - often distorts results in unpredictable ways. The PPCP platform differs from this situation since all translations are pre-validated by professionals. It means PPCP may serve as the backbone for a very large number of uses, including the improvement of machine translation accuracy and validity.

3. Project display

A logged-in user first chooses his/her EHR base language, in which the medical record is written – see the upper right corner of the screen in figure 1. The User interface adapts its format to a variety of devices such as computer screens, tablets and smartphones.

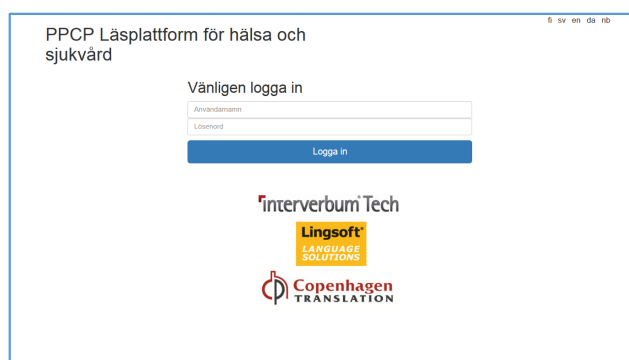


Figure 1: The PPCP login page

Once logged in, the user has a dropdown menu for choosing in which languages the explanations should be provided. The termbase behind the application contains professional and everyday language related to 'heart attack' in Danish, Finnish, Norwegian Bokmål, Swedish and English. Some concepts and terms are also explained in Arabic, Icelandic, Northern Sami, Somali and Tigrinya. The difficult terms, where an explanation is available in the termbase, are highlighted in blue, and the user can see them one by one by pointing the cursor to one of the highlighted words. see figure 2.

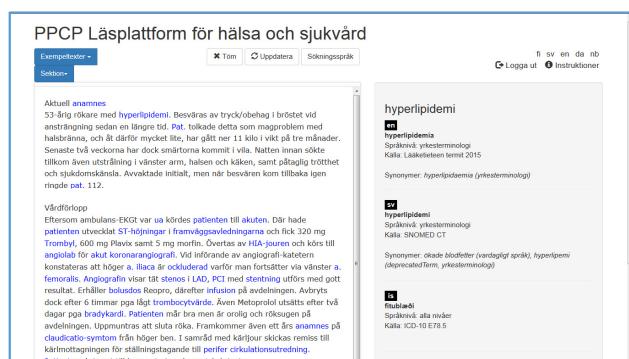


Figure 2: The EHR User interface

Since the content of the medical record is never altered by direct translation or interpretation, the intentions and tone of voice of its authors are never changed. We have learned

this is an important point from a clinical point of view. All terminology and explanations in the database have been developed and validated by specialists. Anyone with a web browser and login information can work in the interface.

The solution is a proof-of-concept to show the principle of how a free-text health record can be made understandable. In addition, multimedia functions can be added to the EHR interface, so that images such as x-rays and magnetic resonance tests related to the patient's status, videos and sound files can be used to further explain concepts and terms. The underlying termbase tool allows this kind of multimedia support.

An online demonstration of PPCP is available on YouTube, <https://youtu.be/PMAH4wS3Xpk>. The demonstration is based on TermWeb, version 3. An updated version using TermWeb 4 will be soon be available.

4. Building the termbase

Since the PPCP project deals with medical terminology, there was a need to identify a relevant clinical subset of terms related to heart attacks. The idea was originally to identify the relevant clinical terms in Snomed CT and then translate them into all relevant languages, with the help of officially mapped results between Snomed CT and ICD-10. This turned out to be more difficult than anticipated due to formal requirements of Snomed International (*SI*, formerly IHTSDO), owner of the Snomed CT database. *SI* does not permit the development of professional terminology into languages of non-*SI* member countries. Consequently, we had to limit ourselves to already existing terms in Snomed CT and their officially mapped equivalents in ICD-10, meaning terms in Danish and Swedish. Existing ICD-10 terminology in other languages were then added. To extend the future usability of our heart-attack-related database, we decided to also include English, since it serves as the base language of both Snomed CT and ICD-10.

As a consequence, patient terms in languages such as Sami, Arabic etc. had to be mapped directly to the professional terms in the main language of the relevant country. Nynorsk was dropped on the advice of Norwegian specialists. Mapping minority language patient terms to the relevant main language concept was a risk mitigation strategy devised already in the original PPCP project plan. After considerable discussions and research, the collection of relevant terms for the database proceeded using the following four methods:

- 1) Introduction of clinical terms subsets (as mentioned above)
- 2) Monolingual term extraction from medical records in the main Nordic languages and English

- 3) Addition of "problem terms" provided by Reference group members such as heart associations
- 4) Supplementation of English term equivalents to Nordic language terms generated using methods 2 and 3

In practice, the work became part of a continuous workflow of gathering, evaluating, translating and validating concepts and terminology of both professional and patient language. The TermWeb tool permits searching for missing entries in the database through specially created filters. An advantage of this iterative process was that it opened for language-to-language translation rather than just professional-to-everyday language translation and vice versa. We also had to choose how to categorise terms, and ultimately came up with compulsory meta-tagging of the 'Language Level' of each term or explanation as either

- All (language) levels
- Everyday language
- Professional language, or
- Jargon and non-standard professional language

While adding terms, great care had to be taken to

- 1) Check that the term did not already exist. This process could be automated by using TermWeb's Batch Search Tool before import, or the Duplicate Check for single-concept editing
- 2) Check that a concept corresponding to the term had not previously been created in the database
- 3) When creating everyday language term equivalents to professional language terms, it sometimes became necessary to enter an explanation rather than a term

Extensive in-house and external resources have gone into developing the heart attack-related database. This includes the use of professional and medically trained translators for Arabic, Somali and Tigrinya. The number of entries for these languages is limited, however, especially in Somali and Tigrinya. The guidelines developed by the PPCP project group for building this kind of medical database can be requested from the authors.

5. Technical solution

The underlying analytical process in PPCP, shown schematically in *Figure 3*, is largely the same as it would be for other potential use cases, enabling further extensions of the platform.

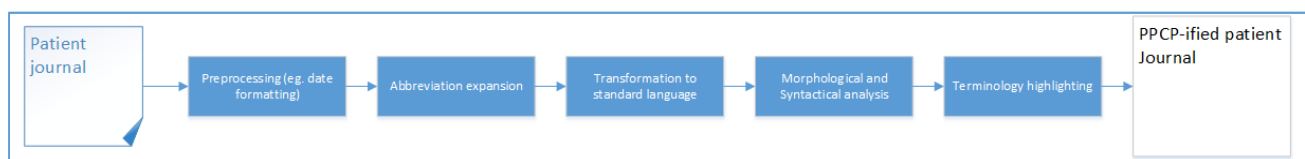


Figure 3: The analysis process for an EHR in the PPCP

A high-level overview of the underlying components in the platform is displayed in figure 4 below.

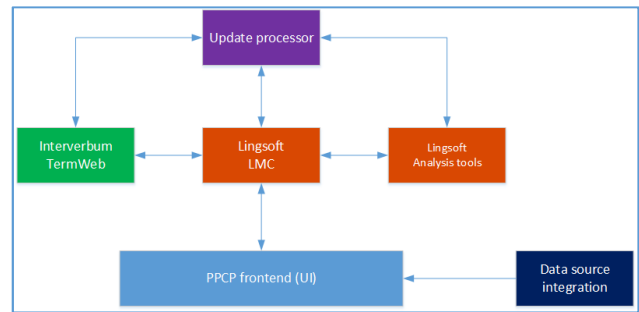


Figure 4: High-level component diagram

The technical infrastructure has been built by using the following existing software:

- Lingsoft's Language Management Central (LMC), a cloud service platform for language analysis. The platform allows authentication and user control for individual users or organisations, and an option for linguistic data collection to allow for analysis of user behaviour and semi-automated updates of the underlying language tools. In PPCP, there is total privacy for any entered information since no logs or copies are kept of the analysed text.
- Lingsoft language tools, the analysis components in LMC, including
 - Word and sentence analysis (entity name recognition)
 - Terminology and semantic analysis
 - Quality checking (spelling, grammar, stylistic rules)
 - Translation
- Intervetum's TermWeb, a terminology management tool, where the multilingual termbase is created and stored

6. Potential fields of application: health sector, eCommerce

- Proactively supporting clinicians in using standardised terminology. By integrating PPCP directly with **Electronic Health Records (EHR)** or writing tools, a physician could continue using his/her regular documentation routines (such as clinic-specific terms) while being prompted for the corresponding standardised terms. It implies dynamic interactive replacement of non-recommended terms. This would, in turn, facilitate automation of reporting requirements for national databases (such as heart or cancer registers) and improve safety aspects of the record. The need for costly terminology training and relearning for active health care providers would be reduced. Reduced costs of training would also be possible for professionals with a different native language.
- When electronic health records use standardised terminology, information in the records may more easily be integrated with clinical decision support systems. These range from regional or national recommendations regarding therapeutic action to the use of broad-based predictive systems such as IBM's Dr. Watson.
- Support for automation of communications between patients and health care providers by text or speech through contact points such as 1177.se, helsenorge.no and sundhed.dk. This communication could be urgent and very specific, where correct triage and documentation is needed.
- Powerful Dr. Watson-type systems that handle Big Data are typically monolingual. PPCP provides a potential avenue for broadening the scope of languages handled.
- Many excellent smartphone and tablet applications for the interactive management of medical conditions lack the 'language plugin' that would facilitate their use among the less educated, old, chronically ill or non-native language speakers. A language plugin would also open new geographic markets for these applications. PPCP can provide the necessary backbone for this multilingual use.

7. Using the termbase for controlled language purposes

In the PPCP interface, terms are recognised and explained in a "free text" context. This function has the potential for wide generic use, for example in e-commerce. PPCP has been demonstrated to individuals involved with Product Information Systems, where product descriptions for e-

commerce platforms are automatically generated from a database. Different e-commerce platforms (Amazon, Ebay etc.) often require different terms for their respective product descriptions. This means "free text" analysis and subsequent conversion of certain terms to what is "correct" for the particular e-commerce platform could be of great interest.

8. Results and conclusions

The main result of the PPCP project is the availability of an accurate and reliable solution for helping patients better understand their own medical records. It is a solution applicable to all languages and language levels, using specialised and easy-to-understand wording. Results have been tested and evaluated by relevant patient groups and clinicians.

Additionally, the PPCP interface allows displaying how a health care professional writing free-text information can be prompted - in real-time - to use standardised rather than non-recommended terms, leveraging the same background functionality and database.

Among the deliverables in the project is also a method for gathering, translating and validating medically accurate terminology in several languages. This method, and the resulting termbase focused on heart attack-related terminology, form part of the project results.

In the PPCP project, only a limited number of languages – those most relevant in the Nordic countries – were included. With adequate semantic recognition support, there is no limit to how many or which languages could be supported by the database. The same generic functionality would also apply regardless of subject domain, given a good terminology database with similar (ISO TBX standard) data categories and values. This opens up a wide range of potential applications where controlled language is used, including e-commerce.

9. Acknowledgements

Nordic Innovation has provided financial support of the project. Special thanks to contributions from individual Reference Group members from several Nordic countries, and to the Danish Heart Association for their terminology support.

From Glossary to Term Base A Tale of Revision – A Question of Quality and Need

Gunnhildur Stefánsdóttir

Translation Centre of the Ministry for Foreign Affairs
Rauðarárstíg 27, 105 Reykjavík
gunnhildur.stefansdottir@utn.stjr.is

Abstract

The Translation Centre of the Ministry for Foreign Affairs has been operated for almost three decades. Its main task is to translate the legal acts pertaining to the EEA Agreement. Integral to the translation environment is the management of terminology in the acts. The term base is one of the largest of its kind in Iceland, with over 80.000 entries at present. A significant factor in the translation is finding and determining Icelandic translations of the specialised terms in the original text. In this paper I describe important aspects of the terminology work carried out in the centre and discuss how this task has developed since the work began. It gradually became clear that the term base needed revision for quality reasons. There have been several phases of revision. I describe how this was performed and the methodology used. It is critical that the term base helps translators to choose and use consistent and correct terminology in the texts. It is widely recognised that quality in translation depends largely on the quality of the terminological resources and their management. To ensure and maintain a level of quality, an ongoing revision and quality control of the term base is necessary.

Keywords: term base revision, terminology management, translation

1. Introduction

The topic of this paper is to give an account of the revision effort of the terminology database that is maintained in the Translation Centre of the Ministry for Foreign Affairs in Iceland. Its main task is the translation the legal acts that fall under the EEA Agreement.

The terminology base is maintained for the purposes of translation, to ensure consistency, harmony and equivalence of terms in the specialised translation environment of the workplace. The selection of terminology is on an ad hoc basis which is a factor that affects how the terminology work is conducted.

The revision effort has been an ongoing activity for several years, although it has not always been conducted in a regular and systematic way. It is worthwhile to consider what reasons and need there have been for reviewing and revising the term entries and information contained in them. An important factor is certainly the requirement of quality in the translations. I will explain how the revision has been conducted in practice and how these methods have been changed and adapted through the years.

2. The Translations and the Terms

The Translation Centre terminology database is the side product of almost thirty years of specialised translation of legal and technical documents. As stated above these are for the most part the legal acts pertaining to the EEA Agreement, but also other documents, such as international agreements, various legal texts, reports, etc., where standardised and specialised terminology is needed.

The terms contained in the term base do not solely originate in the EEA documents, but from the other types of translations as well. Moreover, there are country names, also short standardised phrases that need to be used consistently, names of institutions, acronyms, scientific names in Latin, etc.

2.1 The Workplace

The Translation Centre staff currently counts 34 employees, an administration unit and four translation units each covering several subject areas. There is also a terminology unit, the editor of the term base and three

senior translators/terminology managers, who work part time on the terminology. The main CAT tools used for the translation are SDL Trados Studio and the auxiliary tool MultiTerm. MultiTerm is the inhouse terminology tool used for term look up during translation.

2.2 The Terminology Database

The terminology database of the Translation Centre contains the specialised terminology of the legal acts falling under the Agreement on the European Economic Area; the EEA Agreement, and other specialised texts that are translated by the Translation Centre. The Translation Centre was established in 1990 for the purpose of translating the agreement itself and subsequently the legal acts pertaining to the Agreement. Other translations include international agreements and other legal documents.

Currently, there are over 82.000 entries in the Translation Centre term base. As said before, the terminology created in-house from the translations is stored in MultiTerm term base for in-house translation purposes.

There is also an external publication of validated terms on the web via a computer programme that was designed in the years 2002-2003 and launched on the web in 2003. The web solution does not present any unvalidated entries or any internal notes and information that are visible in the entries in the MT for in-house purposes.

The internet link for the English version of the terminological database on the web is <http://www.hugtakasafn.utn.stjr.is/index.en.adp>. Below in the interface of the Term Base on the internet is shown Figure 1 below.

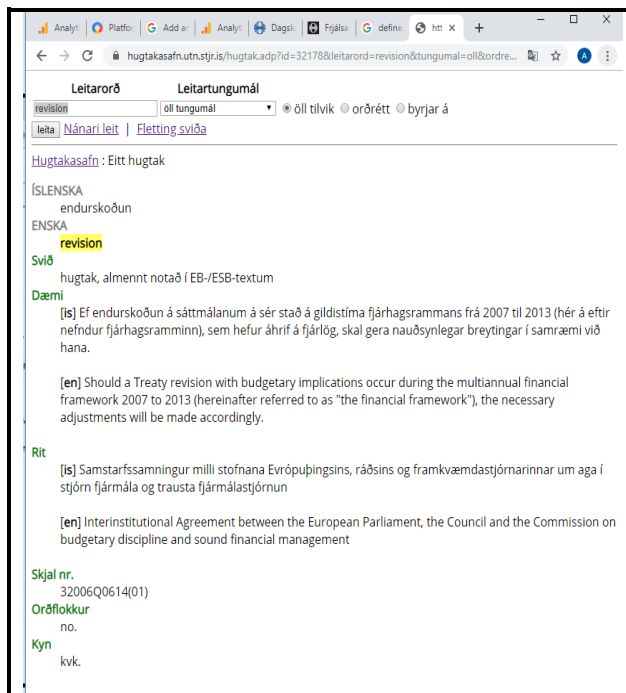


Figure 1: Translation Centre Termbase on the web

It is extensively used by the general public, field experts, students and freelance translators. The frequency of visits to the term base website shows the demand and need for this terminological tool.

3. The Term Base Revisions, Additions, Conversions

There have been various steppingstones and advances in the terminology work and approaches in the thirty years since the terminology compilation began. Below I give a short overview of what I consider to be the most important steps.

3.1 Initial Phases

Terms and vocabulary from the original texts along with the corresponding Icelandic translations were collected in a glossary from the start of the translation unit in 1990 in order to have a reliable source of coherent and uniform translations and use of key terms in the EEA agreement and the subsequent acts pertaining to it. This bilingual glossary of terms was originally maintained in a document in the UNIX system that was used for all text processing at that time. There were three data fields listed in the glossary: the English term, the corresponding Icelandic term and a reference. The reference information was the relevant number of the Official Journal where the legal act from which the term originated was published and the page number where the term appeared in the original text. This glossary grew in volume and constituted the foundation for the terminology data base.

3.1.1 The Bilingual Glossary on the Web

The decision was made in 1995 to upload a version of the bilingual glossary on the Internet. This was quite an innovative action at the time; to use the web in order to make the word list or bilingual glossary available to the general public. At that time, it contained around 15.000 bilingual entries and was uploaded as an alphabetical list

with English as the determinant language alphabetically. Although this was not a very transparent tool in terms of search and look-up, it was well received by translators, specialists, etc., as it contained a lot of LSP (Language for Special Purposes) vocabulary in various areas where there was a gap in Icelandic vocabulary at that time.

3.2 Implementation of Trados and MultiTerm

The first major step in the transition from a bilingual glossary to a terminological database was taken in 2000 when the Translation Centre started using the Trados translation software and the term base programme MultiTerm. This transition entailed a change in the translation environment and the approach to terminology treatment. The bilingual glossary was uploaded to MultiTerm, which then counted approximately 20.000 entries. The entry structure was decided and designed by the editor of the term base and the computer specialist at that time, that is what data categories to use and the field specifications. During this phase and the decisions on the design of the term base, guidance on structuring term entries was sought in the terminology manual by Heidi Suonuutti (1997) and in material from the Icelandic Language Council and consultation with the linguists and terminologists working there.

The searchable fields that were included in MultiTerm were English Term and Icelandic Term and a possibility for multiple addition of terms searchable synonym fields. The fields for English context (text example, Icel. *dæmi*) and Icelandic context (text example, Icel. *dæmi*) were also added but these are not searchable. Other fields that were decided upon and added to the term base were **reference** (Icel. *rit*) in order to enter information on the appropriate legal act, the **celex number** (Icel. *skjalnr.*) of the act, **subject area** (Icel. *svið*) for the title of the chapter/annex of the EEA Agreement, **definition** (Icel. *skilgreining*) in English or Icelandic, **source** (Icel. *heimild*), where the name and title of a field expert who verified the solution should be entered or the title of the source within the subject area, that could be a book, article, web page, etc. After the upload, it became apparent that there were several duplicate entries as well as entries with common vocabulary that had been included in the bilingual glossary. The term base editor carried out a thorough revision, cleaning out various redundant, misleading or erroneous entries. She also worked on the classification of the terms which had not been possible in the glossary version.

3.3 New Tools for Internet Upload and Term Extraction

In 2002 a decision was made to make the terminology contained in-house in MultiTerm available to the general public on the Internet. Along with the programme that was designed for this purpose, a tool for term extraction from the translation texts was designed. These important tools were presented at the Nordterm conference in 2005 and in a paper published in the proceedings of the conference (Sigrún Þorgeirsdóttir, 2006). Both of these programmes constituted a major advance in respect of terminology work, and the web solution greatly improved the access of the public to specialised language. It was possible to systematically search the term base to find information. The term extraction tool worked on the so called BAK files of translated texts that were finalised in Trados and resulted

in a more automatic approach to term extraction in the in-house environment. The creation of new entries increased and, indeed, along with it the need for revision and work on term entries. However, the use of this tool was discontinued when the Trados translation software was upgraded.

3.4 Attempt for a Systematic Overhaul

In 2009 a project was started on a systematic revision of the term base, which at that time counted approximately 40.000 entries. The aim of this project was to go through the entire term base checking each entry in order to verify it.

The plan was for all the translators to take part in the checking, splitting up the term base by subject areas. The reason for this effort was the impending translation of the EU acquis and this was thought to necessitate this review. However, the project was never fully accomplished as the very need for the EU acquis translation that prompted the overhaul led to the discontinuation of this endeavour.

3.5 Steady Increase in Number of Entries – More Focus on Revision

In the period of 2010-2013, during the EU acquis translation, there was a substantive increase of new entries from this translation effort and a huge backlog of unrevised entry proposals was built up.

As from 2010, the revision has gradually become more structured and continuous, and this is in response to the amount of unvalidated proposal entries. A special terminology team was formed, headed by the editor of the term base. From 2017 there has been an ongoing revision project with two full time staff members in a systematic revision of the term base.

4. The Need for Revision

A fundamental question in the context of terminology management is why a revision of a terminology database is needed. The most obvious answer is that it is a matter of quality assurance and is directly related to the reliability and consistency of terminological usage in the translation texts. In the translation environment of legal, technical and standardised texts a correct and updated term base is a critical matter. In the words of Karolina Stefaniak: “Terminology is an integral part of every translation process, necessary to achieve high-quality translation. In the case of EU law, terminology is additionally a matter of legal certainty and clarity. Terminological errors may lead to citizens and companies misunderstanding their rights and obligations” (Stefaniak, 2017).

Another issue regards what such a revision involves and how it is conducted. According to guidelines in a handbook prepared by Vesna Lušicky and Tanja Wissik that is specifically aimed at tackling terminology management in a translation environment, there is an extensive elaboration on appropriate procedures in translation-oriented terminology work (2015).

Lušicky and Wissik explain that they have modified guidelines set forth by Chiocchetti et al. in an article from 2013, saying that “The table in Chiocchetti et al.:2013, p. 31 was adapted to the needs of translation-oriented terminology work” (Lušicky and Wissik, 2015). These guidelines give an excellent overview of what issues and items should be considered when revising a term and are found in Table 1 below:

scope	• minor issue, single terminological entry (can be solved by manually improving single terminological entries or manually creating a single new entry) • recurrent issue, several terminological entries (can be solved by manually improving single terminological entries or batch changes or manually creating a single new entry or batch import)
motivation	• incomplete or incorrect entries • doublettes • legal novelties (e.g. the Treaty of Lisbon amending the Treaty on European Union and the Treaty establishing the European Community) • innovation in subject fields (e.g. obsolete terms, new emerging terms) • enlargement of the termbase (e.g. adding additional subject fields) • language reforms (e.g. spelling reforms)
frequency	• daily • periodically (e.g. once a month) • translation-oriented (e.g. after having finished the translation of certain documents/ a certain field)
manner	• manually • (semi-)automatically

Table 1: Items to consider in terminology revision

4.1 Periodic Revision

For several years the revision of the terminology was not conducted on a regular basis. The main emphasis was on increasing the number of entries, and this is understandable in view of the need of extensive terminology for translation purposes. If errors were found, they were quickly and carefully corrected but there was not a systematic process in place to review or revise the term base as a whole. Nonetheless, always when the opportunity arose, the term base was purged, for instance with the upload from the UNIX-based glossary into MultiTerm when there turned out to be several duplicate entries. They were all deleted and dubious entries looked into and amended or deleted.

4.2 Gradual Change of Methods

There were several aspects, both internal and external, that made us increasingly aware of the need for a more consistent and regular revision of the term base. To name the most influential ones, there was more contact and cooperation with the Icelandic Language Council and the terminology committees operating partly under their auspices. Also, the Translation Centre began participation in the activities of both Nordic and European terminology groups and associations. One of these is the Nordterm terminology association and the other is the Conference of Translation Services of European States (COTSOES). A great deal of knowledge on terminology management was gained through participation with these entities, and also on the technical aspect of managing termbases.

Another factor was the increased cooperation with experts in various specialised fields. Information and details from them provided us with a sharper perspective and more knowledge on the terminology in the field in question and made it possible to amend and correct term entry information.

Yet another reason was the launching of the IATE terminology database on the internet. This tool became a standard look up site for the translators, revisors and the terminologist. Substantial knowledge was sought from this

termbase and many of its features became a model for the working methods in the Translation Centre.

Lastly, a mention should be made of the ISO standards on terminology, primarily the standard on terminology work - principles and methods (ISO 704, 2009) which made a substantial impact on the attitude towards terminology management and how it should be conducted. These were purchased in 2009 and looked into, and even if they were not adopted for use in the translation environment, the content was used, and served to improve the in-house methods and approaches to terminology management.

4.3 Current Revision

The revision of the terminology base gradually became more structured, regular and continuous. It also became more systematic in terms of what issues were to be checked in particular. There are now in place well-established standard procedures on how to conduct revision of the terminology base. These have been developed in-house by the editor of the term base and the terminology team.

The inhouse guidelines for the current revision project entail a checking procedure. Below is a translation of the inhouse document for the procedure. A list is retrieved of all entries in a specific subject area. Each term entry is checked, and the following procedure carried out:

- if outdated, obsolete, redundant, etc the term is deleted from MultiTerm
- a copy is pasted in a special document for outdated entries
- if a change is made it is noted in the note field: "former translation was ..., translation changed YEAR"
- outdated example replaced with a new one from the EEA text corpus and corresponding change made in the reference field (Title of EU act)
- correct celex number entered in the relevant field
- check if source is sufficiently entered, fx year missing, part of title
- look up in IATE/Icelandic Word Bank for synonyms, deprecated terms, definitions, etc
- look up related terms (multi word) and make necessary changes and notes

In this process, there emerges a substantial number of entries that have not been edited or revised, and lack information, such as definition, context, etc. However, it also turns out in the checking, that the majority of the terms in the term base has been partially or fully reviewed and revised.

4.4 Revision Issues and Examples of Entries

As has been mentioned earlier, there are several reasons for conducting a regular and methodical revision of terminology bases. The most important one would be the urgent requirement of quality of the terminological content for translation purposes, especially when translating legal, technical and specialised language content. The significance of quality assurance has been addressed in a number of publications on terminology management. In a recent article on quality assurance in multilingual legal terminological databases, Chiocchetti et.al. state that "In addition to several other factors, terminology and terminology management are perceived to be important elements of ... translation ... and multilingual law-making. This implies that quality in translation, including legal

translation, partly depends on the quality of terminology and terminology management." (Chiocchetti et.al, 2017). Another factor should also be mentioned: a high quality and reliable term base facilitates the translation work because the translators are able to rely with more confidence on the terms and the information there. This goes hand in hand with the necessity to manage terminology well with the required revision.

4.4.1 Revision Issues

In informal personal communication with several in-house translators who also work on terminology, I found that in their opinion the most important information in an entry is the categorisation of the term and that this should be accurate. Other significant information is to have a definition of the term. Also, they considered it very helpful to have clear context, that is usage examples in both English and Icelandic. Some also found it helpful to see the source information, for instance that the translation term came from a certain field expert. Yet another type of information in the entry was deemed to be helpful, this was the term in the Scandinavian languages, or German or French. This was in the case of very nebulous meaning of the term in English.

All this contributes to the way in which the translator makes his or her decision on what term to use or whether more research is needed. This is also a question of how quickly this decision can be made in the translation process.

4.4.2 Examples of Entries

In order to show examples of what the task of revision entails, there are three pictures of entries, that are in different stages of the revision process. In Figure 2, an example is shown of an entry that has not yet been revised. Context, definition and reference (the title of the legal act) is missing.

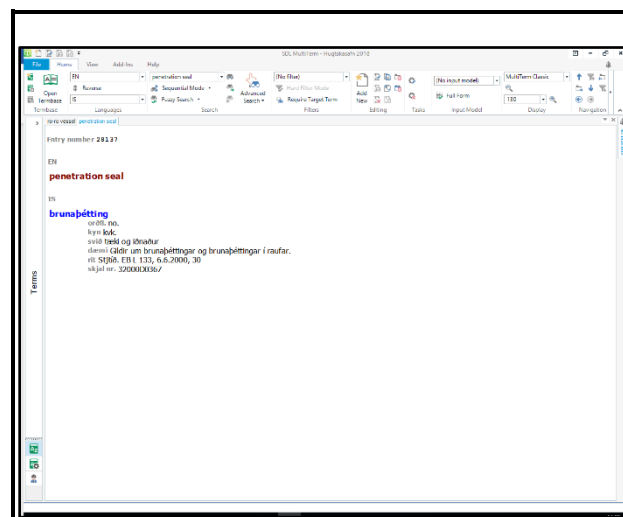


Figure 2: An unrevised entry

Figure 3, below, shows an entry that has been partially revised, context in English and Icelandic has been added and the name of the field expert has been entered in the source field.

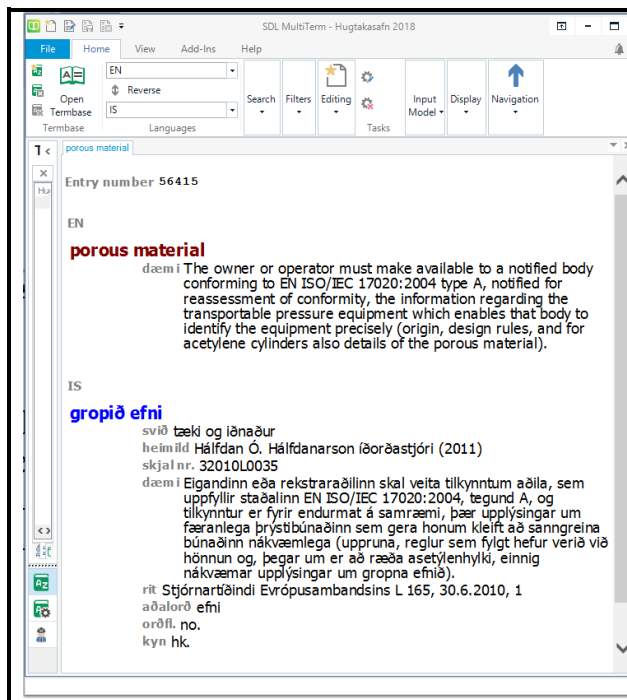


Figure 3: A partially revised entry

The entry shown in Figure 4 has been fully revised. There is both English and Icelandic context, definition in both English and Icelandic, the reference in both the languages, guidance on usage, the source is listed, and the term is also given in Danish, Swedish, German and French.

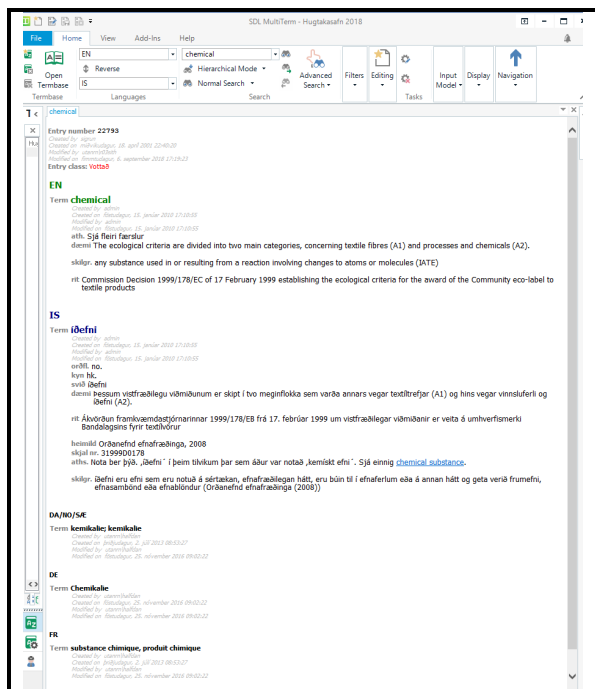


Figure 4: A fully revised entry

5. Concluding remarks

There has been an ongoing compilation of terminology for nearly three decades in the Translation Centre of the Ministry for Foreign Affairs in Iceland. This is specialised terminology originating for the most part in the legal acts falling under the EEA Agreement and other legal and technical material translated in the Centre. Currently, there

are around 82.000 entries in the term base and it constitutes one of the largest terminological resources in Iceland.

Through the years, the emphasis on terminology management and quality assurance of the terms contained in the term base has steadily increased. The importance of a high quality and reliable terminology for the purposes of accuracy and consistency in the translation work has always been clear. The focus is therefore on regular and ongoing revision of the terminological content. Currently, there are in place standard procedures on how to conduct the revision and a special revision project is being conducted.

There have been gradual changes in the working environment, for instance technical advances that have influenced the approaches to terminology work and how it is conducted. Cooperation and participation in professional terminology groups also served to increase the standard of our work through the knowledge thus attained.

The term base has been maintained in-house in the term base MultiTerm and is integral tool in translation process. It is also accessible on the internet, but that version contains only validated terms.

This has been an overview of the gradual process of modifications in the terminology management in the Translation Centre. This has come about from both internal and external forces. The central issue is the importance of quality and what measures can be taken to maintain it.

6. Bibliographical References

- Chiocchetti, E., Wissik, T., Lušický, V., & Wetzel, M. (2017). Quality assurance in multilingual legal terminological databases. *The Journal of Specialised Translations*, 27, pp. 164-188
- COTSOES — Conference of Translation Services of European States Working Party on Terminology and Documentation (ed.) (2002). *Recommendation for Terminology Work*. Bern: Federal Chancellery.
- COTSOES — Working Group on Terminology and Documentation. (ed.) (2016) *Terminology Management for Translation: A Starter Guide*. Unpublished draft. Retrieved from Translation Centre Intranet 28 September 2019.
- Lušický, V. and Wissik, T. (2015). *Procedural Manual on Terminology*. Translation-Oriented Terminology Work. Published in cooperation with the Sector for Preparation of the National Version of Acquis Communautaire, Secretariat for European Affairs <http://www.sep.gov.mk/data/file/Preveduvanje/Procedural Manual on Terminology final version.pdf>
- Stefaniak, K. (2017). Terminology work in the European Commission: Ensuring high-quality translation in a multilingual environment. In T. Svoboda, Ł. Biel & K. Łoboda (eds.), *Quality aspects in institutional translation*. Berlin: Language Science Press, pp. 109–121. DOI:10.5281/zenodo.1048192
- Suonuuti, H. (1997). *Guide to terminology*. Nordterm 8, Helsingfors: Tekniikan Sanastokeskus ry/Nordterm Translation Centre. (2019). Guidance on terminological revision. Retrieved from Translation Centre Intranet 28 September 2019.
- Þorgeirsdóttir, S. (2006). Islands Udenrigsministeriums Oversættelsescenter: Ny termdatabase og nye metoder ved indsamling af termer. In Ágústa Þorbergsdóttir (ed.), Nordterm 14, *Nordterm 2005: Ord og termer*:

proceedings fra Nordterm 2005. Reykjavík: Íslensk málstöð, pp. 153-167

7. Language Resource References

Hugtakasafn þýðingamiðstöðvar utanríkisráðuneytisins
<https://hugtakasafn.utn.stjr.is/umhts.adp>

IATE (InterActive Terminology for Europe):
<http://iate.europa.eu/SearchByQueryLoad.do?method=load>

ISO 704 (2009). ISO 704: Terminology Work -- Principles and Methods. ISO Technical Committee 37 (ed.). Switzerland: International Organisation for Standardization, ISO.

Relationer i Skatteforvaltningens Informationsmodeller

Hanne Erdman Thomsen

Informationsarkitekt

Skatteforvaltningen, Udviklings- og Forenklingsstyrelsen

Østbanegade 123, København

hanneerdman@gmail.dk

Abstract

Skatteforvaltningens informationsmodeller i UML indeholder klasser med attributter og relationer mellem klasserne. Relationerne omfatter specialisering, som svarer til typerelationen, og uspecificerede relationer, som navngives ad hoc. På Handelshøjskolen CBS er der tidligere arbejdet med at opstille et inventar af relationer til brug i terminologiarbejde, og dette inventar har siden været anvendt i bl.a. Kriminalforsorgen. I dette indlæg vil jeg anvende relationsinventaret på eksempler fra Skatteforvaltningens informationsmodeller, og se om der er relationer i eksemplerne som ikke er dækket ind i inventaret.

Keywords: relationer, begrebsmodeller, informationsmodeller

1. Udvidet Abstract

1.1 Semantiske relationer i ontologier

I forbindelse med projektet OntoQuery (Andreasen et al. 2002) arbejdede vi med semantiske relationer i ontologier (Madsen et al. 2002). I forlængelse af dette arbejde udarbejdede DANTERMcentret i samarbejde med Sundhedsdatastyrelsen et samlet inventar på 24 relationer omfattende typerelationen, fire forskellige delhelhedsrelationer, to temporale relationer og 17 associative relationer (Sundhedsdatastyrelsen (intet årstal)). De associative relationer er karakteriseret ved hvilke typer begreber de forbinder: aktiviteter, roller, lokationer samt den meget generelle type, entitet. Der er for eksempel 4 aktivitet-rolle-relationer: aktivitet-agent, aktivitet-patient, aktivitet-resultat og aktivitet-instrument. Alle relationerne navngives med et verbum (evt. med præposition) i hver retning. Således kaldes aktivitet-instrument-relationen henholdsvis *udføres ved hjælp af* og *anvendes ved*:

Brændehugning udføres ved hjælp af en økse

En økse anvendes ved brændehugning

DANTERMcentret har desuden anvendt relationsinventaret i forbindelse med praktisk begrebsafklaringsarbejde hos blandt andet Kriminalforsorgen i Danmark (Madsen, B.N. 2012).

1.2 Skatteforvaltningens informationsmodeller

I Skatteforvaltningen udarbejdes der informationsmodeller i forbindelse med it-udviklingsarbejde, og i disse modeller navngives relationer ad hoc, uden en systematisk tilgang. Der er ikke tale om terminologiske begrebsmodeller, som de terminologiske begrebsrelationer er beregnet på, bl.a. er klasserne forsynet med attributter og der tilføjes kardinalitet på relationer.

1.3 Ontologiske relationer i informationsmodeller

I mit indlæg vil jeg vise i hvilket omfang det er muligt at erstatte ad hoc-relationerne i Skatteforvaltningens informationsmodeller med de standardiserede relationer som anvendes i begrebsarbejdet hos Kriminalforsorgen.

2. Referencer

Andreasen, T., Jensen P.A., Fischer Nilsson, J., Paggio, P., Sandford Pedersen, B., and Thomsen H.E. (2002). OntoQuery: Ontology-based Querying of Texts. Trykt i: *Acquiring (and Using) Linguistic (and World) Knowledge for Information Access*. Jussi Karlgren, Pentti Kanerva & Björn Gambäck (ed.). Working Notes. March 25 - 27 2002, Stanford University. AAAI-2002 Spring Symposium Series, pp 28-31, Stanford.

Madsen, B.N. (2012): *Vejledning for begrebsarbejde i Kriminalforsorgen*. Kriminalforsorgen og DANTERMcentret. Hentet 28-02-2019 på <http://kriminalforsorgen.terms.dk/search/filer/H%C3%A5ndbog%20i%20begrebsarbejde%202012-09-18.pdf>

Madsen, B.N., Sandford Pedersen, B., and Thomsen H.E. (2002). Semantic Relations in Content-based Querying Systems: a Research Presentation from the OntoQuery Project. Trykt i: Simov, K. & Kiryakov, A. (ed.): *Ontologies and Lexical Knowledge Bases: 1. International Workshop, OntoLex 2000, Sozopol, Bulgaria, September 8-10, 2000 – Proceedings*. Sofia ; Onto Text Lab.

3. Referencer til sproressourcer

Sundhedsdatastyrelsen. (intet årstal). Begrebsbasen. Tilgået 28-02-2019 på <https://sundhedsdatastyrelsen.dk/da/rammer-og-retningslinjer/om-terminologi/nbs>.

Nye perspektiver og udfordringer for terminologi i en islandsk kontekst

Ágústa Þorbergsdóttir

Árni Magnússon-instituttet for islandske studier
Reykjavík, Island
agusta.thorbergisdottir@arnastofnun.is

Abstrakt

Som bekendt er de såkaldte termudvalg en udbredt ordning i terminologiarbejdet i Island. Det er oftest fageksperter som står bag tilblivelsen af termlister og andet fagsprogsarbejde. Drivkraften for deres arbejde er i de fleste tilfælde baseret på deres interesse, og de deltager i et frivilligt arbejde. Blandt andet af denne grund kan termarbejdet være lidt tilfældigt, og det viser sig, at i den islandske termbank, Íðorðabankinn, hvor de fleste termlister er publiceret, er det nemt at finde domæner, hvor der mangler termlister.

Med projektet eTranslation TermBank (eTTB), som har til opgave at indsamle terminologiske ressourcer til at forbedre kvaliteten af EU's automatiske oversættelsessystem, er der lagt en ny kurs i islandsk terminologi. Aldrig før er der blevet gjort en sådan målrettet indsats for at indsamle terminologiske ressourcer i Island. I mit foredrag vil jeg diskutere resultaterne af projektet. Jeg vil argumentere for vigtigheden af denne indsamling af terminologiske ressourcer (eTTB) for små sprogsamfund som det islandske, og som forhåbentligt kan anvendes videre i udvikling af sprogteknologiske værktøjer.

Keywords: terminologiarbejde, termbank, sprogteknologi

1. Indledning

Denne artikel præsenterer nye perspektiver og udfordringer for terminologi i en islandsk kontekst. I begyndelsen gøres der rede for hvad der hidtil har været traditionelt termarbejde i Island, og hvilke fordele og ulemper der har været ved denne ordning. I nutidens digitale tidsalder, hvor der også bør tages højde for, hvordan man bedst kan sikre brugen af islandsk i digitale tjenester, er det traditionelle arbejde tydeligvis ikke længere tilstrækkeligt. EU's strategi for det digitale indre marked, hvor informationer på tværs af grænserne skal være mulige uden sproglige barrierer, stiller krav om velfungerende automatisk oversættelse.¹ I denne sammenhang overvejes den sprogteknologiske situation for islandsk samt erfaringerne af projektet eTranslation TermBank og dets værdi for islandsk.

2. Traditionelt termarbejde i Island

Termarbejde udført af termudvalg og individer er en udbredt ordning i Island, men myndighederne har også ydet vigtigt bidrag til termarbejde i Island, ikke mindst med Termbanken.

2.1 Termudvalg

Initiativ til at udføre terminologisk arbejde i Island tages først og fremmest af de såkaldte termudvalg. Over 50 termudvalg har eksisteret i Island i en længere eller kortere tid. Det er en gammel tradition som daterer sig tilbage til 1919 da det første udvalg blev oprettet af Ingeniørforeningen.

I 1980'erne og især 1990'erne tog terminologisk aktivitet i Island et stort spring fremad, og antallet af termudvalg steg (Dóra Hafsteinsdóttir 1998:51). Der kan være tale om en række forskellige årsager. Her kan især nævnes

Termbankens oprettelse i 1997² og forberedelsen til to store projekter i 1990'erne, dvs. engelsk-islandsk ordbog og en islandsk encyklopædi. Arbejdet på disse bøger krævede ensartede islandske termer, som specialister inden for et bestemt fag kunne acceptere. Det var ret almindeligt, at der manglede islandske termer, og ofte var der mange forskellige termer for ét begreb (Dóra Hafsteinsdóttir og Sigríður Harðardóttir 2001:19). Dette problem måtte løses. Fageksperter blev konsulteret, og det blev forsøgt at harmonisere og finde én term, som alle kunne være enige om. I forlængelse heraf blev flere termudvalg oprettet og flere termsamlinger publiceret.³

Udvalgene er af forskellig slags, bl.a. hvad angår aktivitet og arbejdsmetoder. De har dog til fælles, at der er tale om en gruppe individer – sædvanligvis fageksperter – som tager sig af terminologien inden for deres fag og udarbejder fagordbøger og termsamlinger. Udvalgene kan udnævnes af fagforeninger eller institutioner, men undertiden er det en uformel gruppe af eksperter som bliver til f.eks. omkring et projekt. Det er hellere ikke ualmindeligt, at et initiativ til termarbejde kommer fra individer, f.eks. fra universitetslærere eller oversættere, som har lavet termsamlinger i forbindelse med deres arbejde (jfr. Sigrún Helgadóttir og Kristján Árnason 1991:8–10).

2.2 Myndighederne

Da de ældste udvalg blev oprettet, havde islændingene ingen institutioner, der kunne tage sig af dette arbejde. Initiativet måtte derfor komme fra individer, og man kan derfor sige, at termudvalgene og individerne har lagt kursen, hvad angår termarbejde (jfr. Sigurður Jónsson 2007:162).

Myndighedernes involvering af et såkaldt neologiarbejde kan føres tilbage til halvtredserne, men da uddelte undervisningsministeriet midler til dette arbejde. Det var

¹https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-ettranslation_en.

²Daværende navn var *Orðabanki Íslenskrar málstöðvar* (Islandsk sprogsekretariats ordbank), <http://ismal.hi.is/ob>.

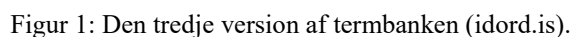
³Arkitekternes termudvalg kan nævnes som et eksempel på termudvalg som blev oprettet i forbindelse med den islandske

encyklopædi <https://idord.arnastofnun.is/leit/ordabok/ARKITEKT> og en termsamling i meteorologi som blev lavet i forbindelse med engelsk-islandsk ordbog <https://idord.arnastofnun.is/leit/ordabok/VO>.

Det var imidlertid først efter midten af det 20. århundrede, at der fandtes institutioner, som kunne påtage sig termarbejde. Islandsk sprogævn (Íslensk málnefnd) blev oprettet i 1964, men det største skridt mod terminologiarbejde blev dog taget, da Islandsk sprogsekretariat (Íslensk málstöð) blev oprettet i 1985. En af dets opgaver var at have samarbejde med termudvalg og andre, som arbejder med deres fagspecifikke terminologi, og disse ændringer gav nyt liv til terminologiarbejdet i Island (Sigurður Jónsson 2007:25–26, Halldór Halldórsson og Baldur Jónsson 1993:123–124).

Tiderne har ændret sig. Det er ikke tilstrækkeligt, at de terminologiske ressourcer kun findes i Termbanken eller i bogform. I den digitale tidsalder må der tages højde for andre faktorer vedrørende terminologi. Hvis det islandske sprog skal overleve i den digitale tidsalder, kræver det nye metoder, og der skal investeres i islandsk sprogteknologi.

I 2012 kom Meta-Net-rapporten, som kortlagde den sprogteknologiske situation for 31 forskellige europæiske sprog som led i EU-projektet Meta-Net. Alle sprogene blev undersøgt på baggrund af ensartede parametre inden for fire områder: tekstanalyse, sprog- og tekstresurser, taleteknologi og maskinoversættelse. Islandsk lå overalt i den dårligste kategori (ringe eller ingen støtte) på den europæiske liste for vurdering af det sprogteknologiske niveau. (Eiríkur Rögnvaldsson et al. 2012:26–30).



I 2006 blev Islandsk sprogsekretariat lagt sammen med fire andre institutioner i Arni Magnússon-instituttet for islandske studier. Af bemærkningerne til den lov, som nu er vedtaget om instituttet (nr. 40/2006), fremgår at det skal varetage forskellige terminologiske opgaver bl.a. rådgivning og vejledning om terminologiske arbejdsmetoder, overtage driften af termbanken, samt have et samarbejde med termudvalg og andre som arbejder med deres fagspecifikke terminologi. Disse bemærkninger bygger på ældre love og bekendtgørelser om Islandsk sprognævn og Islandsk sprogsekretariat.⁴

Afburðagóður stuðningur	Góður stuðningur	Sæmilegur stuðningur	Brotakenndur stuðningur	Litill/enginn stuðningur
	Enska	Franska Hollenska Íralska Pólska Spænska Sænska Tékkneska Ungverska Þýska	Baskneska Búlgarska Danska Eistneska Finnsla Galisíska Gríska Katalónska Króatíska Norska Portúgalska Rúmenska Serbneska Slóvakíska Slóvenska	Írsla Íslenska Lettneska Litháíska Maltneska

Figur 2: Islandsk lå i den dårligste kategori i Meta-Net rapporten.

Til alt held er situationen blevet bedre siden 2012 da Meta-Net rapporten blev puliceret. I de seneste år er der blevet arbejdet strategisk med sprogteknologi for det islandske sprog. Der har været en række projekter og programmer med det formål at oprette tekstresurser og især ordbogsresurser for islandsk. Island har vedtaget en strategi for udvikling af sprogteknologi for islandsk, *Måltækni fyrir*

Jeg har nu i korte træk beskrevet det traditionelle termarbejde i Island. Det har sine fordele og ulemper. Til en vis grad er det gået ganske godt. I de senere årtier er der opbygget betydelige ressourcer i Termbanken, og den har stærke domæner, især medicin. Hvis vi ser på ulemperne, så mangler der termsamlinger på diverse fagområder i Termbanken. Det er en svaghed og følge af frivilligt arbejde i termudvalgene. Der findes ikke en overordnet

⁴Lov nr. 80/1984, lov nr. 2/1990, lov nr. 40/2006. Bekendtgørelse nr. 49/1965, bekendtgørelse nr. 159/1987.

islensku 2018–2022, som nu er blevet implementeret.⁵ Der findes både en islandsk sprogteknologifond, *Máltækni sjóður*, der har til opgave at finansiere projekter, som udvikler sprogteknologi for islandsk⁶, og et strategisk forskningsprogram for islandsk sprogteknologi, *Markáætlun í tungu og tækni*.⁷

Relativt få mennesker taler islandsk, eller ca. 330.000, og i markedsmæssig forstand er det ikke interessant for udenlandske producenter at tilbyde islandsk i deres sprogteknologiske produkter. Hvis det islandske sprogs fremtid skal sikres, er det nødvendigt med offentlig støtte til udvikling af sprogteknologi. Denne udvikling kræver adgang til store mængder af sproglige data, og de terminologiske ressourcer må ikke glemmes i denne sammenhæng. Digitale sprogressourcer på islandsk er dårlige. Jo flere data der bliver offentligt tilgængelige, desto bedre sprogteknologi kan der udvikles.

4. eTranslation TermBank

Med deltagelse i projektet eTranslation Termbank er der lagt en ny kurs i islandsk terminologi. Det er det første skridt til sprogteknologiske anvendelser i islandsk terminologi.⁸

Formålet med projektet var at indsamle terminologiske ressourcer inden for specifikke fagdomæner for at forbedre kvaliteten og dækningsgraden af EU's oversættelsessystem CEF eTranslation. Fokus var sat på domæner knyttet til følgende online-tjenester: e-Justice (retsinformation), eHealth (sundhed) og ODR Open Dispute Resolution (forbrugerbeskyttelse).

Projektet gav en god anledning til at indsamle terminologiske ressourcer, og samarbejdet med de syv andre lande var lærerigt for Island. Aldrig før er der blevet foretaget en så målrettet indsamling af terminologiske ressourcer i Island. Her var termlisterne fra Termbanken ikke nok. Der var et stort behov for at skaffe flere data.

4.1 Indsamlingen

I Island var Árni Magnússon-instituttet for islandske studier ansvarlig for at indsamle og uploade parallelle tekster og terminologiske ordlister til projektet. Hovedsageligt henvendte man sig til offentlige institutioner, som ifølge lov om videreanvendelse af offentlige informationer nr. 45/2018 skal levere deres data til gavn for samfundet som helhed.

Hagstofan (Statistics Iceland)
Landlæknir (Directorate of Health)
Persónuvernd (The Icelandic Data Protection Authority))
Neytendastofa (Consumer Agency)
Ríkiskaup (State Trading Centre)
Ríkislögreglustjóri (National Commissioner of the Icelandic Police)
Ríkisskattstjóri (Directorate of Internal Revenue)
Seðlabanki Íslands (Central Bank of Iceland)
Þýðingamiðstöð utanríkisráðuneytisins (Translation Centre, Ministry for Foreign Affairs)

Figur 3: Eksempler på indsamling fra offentlige institutioner

I opsporingen af de terminologiske ressourcer henvendte man sig også til andre, f.eks. private institutioner hvis man havde en forventning om at de skulle råde over terminologi. Der kan f.eks. nævnes at afdelingen for Sygeplejevidenskab på Islands Universitet hjalp til med at skaffe påkrævet brugertilladelse for ICNP – International Classification for Nursing Practice såvel på islandsk som på engelsk.

I den islandske termbank fandtes også termsamlinger som gavne projektet eTranslation Termbank. Af de tre domæner, hvor terminologiske ressourcer skulle indsamles, var eHealth det bedst dækkede domæne i den islandske termbank. Her skal dog nævnes, at inden for eHealth eller sundhed er der forskellige termsamlinger som utvetydigt hører under sundhedsdomænet, f.eks. medicin, epidemiologi, biologi, farmakologi, immunologi. De andre to domæner, forbrugerbeskyttelse og retsinformation, er i mindre grad dækket i Termbanken.

4.1.1 Erfaringer

Hvordan gik så det første sprogteknologiske projekt i islandsk terminologi? Var det nemt at indsamle ressourcerne?

Almindeligvis gør de offentlige islandske institutioner ikke meget ud af terminologiarbejde, og det var ikke alle, som havde termlister eller parallelle tekster, som de kunne levere. Projektet viste sig at være kompliceret for mange. De nåede ikke frem til sagens kerne, og de så ikke ud til at være opmærksomme på værdien af deres sprogressourcer for udviklingen i islandsk sprogteknologi.

Hele forløbet omkring indsamlingen var tidskrævende. Det krævede gentagne henvendelser, ofte mange telefonsamtaler, en række e-mails og undertiden personlige møder at få leveret terminologi. For at opnå resultater behøvedes der ihærdighed, men med tiden blev forløbet smidigere, dvs. man lærte at udnytte erfaringerne, f.eks. hvem man skal henvende sig til.

Licensbetingelser, f.eks. Creative Commons systemet, var forvirrende for nogle, som ikke forstod formålet med det og behovet for åbne data. Der fandtes eksempler på, at

⁵Máltækni fyrir íslensku: <https://www.stjornarradid.is/lisalib/getfile.aspx?itemid=56f6368e-54f0-11e7-941a-005056bc530c>.

⁶Máltækni sjóður: <https://www.rannis.is/sjodir/rannsoknir/maltaeknisjodur>.

⁷Markáætlun í tungu og tækni: <https://www.rannis.is/sjodir/rannsoknir/markaetlun-i-tungu-og-tækni>.

⁸eTranslation Termbank : <https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/2016-eu-ia-0122>

besvær med brugertilladelsen forsinkede indlevering af ressourcerne, f.eks. at glemme at registrere et antal lemmaer i pågældende ressourcer eller at underskrive brugertilladelsen.

Angående de parallelle termlister fandtes i det mindste et eksempel på tilladelse til at bruge den islandske version, men ikke den modsvarende engelske version. (ICD-10 klassifikation af sygdomme (diagnoseliste) udformet af Verdenssundhedsorganisationen. Dette er eksempel på hvor uheldigt det kan være, når ressourcer af høj kvalitet er licensbelagte og derfor fravælges.

4.1.2 Forhindringer med ressourcerne

Man kunne ofte mærke, at ikke alle var parate at give ressourcerne fra sig. Der var mange argumenter. Det var ikke ualmindeligt, at det blev nævnt, at ressourcerne ikke var færdige nok til aflevering. Det skulle være nødvendigt at læse det hele grundigt igennem. Det forsinkede indsamlingen.

Der var også bekymring over eventuel opdatering af ressourcerne senere, dvs. at opdateringen ikke sker alle steder. Det indsamlede materiale bliver forældet, hvis der ikke sker en løbende indsamling af ressourcer. Det var som om folk frygtede at miste grebet om ressourcerne, som kunne blive forældrede. Desuden var det svært at forstå, hvor terminologien lagres, og hvordan den bliver tilgængelig for almenheden.

Ressourcerne var ikke altid i et passende format, dvs. ikke i et struktureret og maskinlæsbart format, og det kunne være svært at håndtere dem. Engelsk og islandsk passede ikke altid godt sammen, f.eks. fandtes der ressourcer med to spalter med engelske ord men én spalte med islandske ord. Oversættelser var hellere ikke altid ækvivalente, f.eks. engelsk „cleaning assistant“ (rengøringsassistent) kan næppe modsvare „störf við ræstingar“ (rengøringsarbejde).⁹

4.1.3 Udfordringer

Den islandske sprogpolitik, *Íslenska til alls* (Islandsk til alt), har, som titlen antyder, som overordnet målsætning, at islandsk bliver brugt inden for alle samfundsområder.

I Island bruges engelsk meget i stedet for islandsk (jfr. Birna Arnbjörnsdóttir og Hafdís Ingvarsdóttir 2018). I digitale sammenhænge er det f.eks. almindeligt at engelsk anvendes til søgning på Google og Wikipedia. Dette er en udfordring for udviklingen af sprogteknologi.

Indsatsen med eTTB-projektet inden for de tidligere omtalte digitale tjenester kan forhåbentlig forhindre, at denne digitale tjeneste kun bliver indført på engelsk. Ikke alle behersker fuldstændigt engelsk, og det er til gavn for alle at kunne bruge deres modersmål. Foruden at forbedre kvaliteten af EU's oversættelsessystem er det ikke mindre vigtigt, at det kan bidrage til at overvinde sprogbarrierne i den digitale verden, hvor det engelske sprog er

dominerende og dermed forbedre dagliglivet for almene europæiske borgere.

Det er vigtigt, at modersmålet kan fungere i alle situationer i det digitaliserede informationssamfund. Det kan endda være udslagsgivende for det islandske sprogs overlevelse. Hvis vi ikke kan holde liv i modersmålet og bruge det i det moderne informationssamfund, har sproget en lav status og bliver kun til hjemmebrug. Et sådant sprogtab er også tab af kultur.

5. Konklusion

Det traditionelle islandske termarbejde med termudvalgene er utvivlsomt blevet til gavn for det islandske sprog, og der er ikke noget som indikerer at de er på vej ud. Termbanken har også haft en positiv indvirkning, og med den blev det muligt at tilgængeliggøre termsamlinger, til stor nytte for mange.

Tiderne har dog ændret sig, og en termbank er ikke længere tilstrækkelig. Udviklingen i informationsteknologien er eksplosiv og kan være en trussel for det islandske sprogs fremtid. Der er brug for en ny kurs, hvis islandsk skal overleve i den digitale verden, og det er vigtigt at udvikle islandsk sprogteknologi og indsamle sprogressourcer.

Styrkelse af sprogteknologi kan føre til større effektivitet i samfundet og gavne den enkelte borger. Hvis alle EU-borgere uanset sprog kan bruge deres modersmål i de digitale tjenester, så er der tale om liget, hvor sprogbarrierne bekæmpes.¹⁰

Indsamlingen af de terminologiske ressourcer for eTTB-projektet var krævende, men ikke desto mindre betydningsfuld for Island. En god del af termsamlingerne i termbanken udnyttedes i projektet, især inden for sundhedsdomænet (eHealth).

Med eTTB-projektet har et stort antal mennesker i Island fået præsenteret vigtigheden af indsamling af sprogressourcer, og det kan forventes, at det har ført til større bevågenhed om værdien af termlister og tekster for udviklingen af sprogteknologi som en væsentlig faktor til beskyttelse af sproget.

Der er en stigende interesse for at lægge øget vægt på sprogteknologi i det islandske terminologiarbejde, og eTTB-projektet var en god begyndelse. Parallelt med de traditionelle termudvalg og deres arbejde er det ønskværdigt at fremme og udvikle islandsk sprogteknologi og gribe de muligheder, hvor sprogteknologi kan bidrage til terminologien.

6. Litteratur

Anna Björk Nikulásdóttir, Jón Guðnason og Steinþór Steingrímsson. (2017). *Málæðni fyrir íslensku 2018–2022: Verkætlan*. Reykjavík: Mennta- og menningarmálaráðuneytið.

<<https://www.stjornarradid.is/lisalib/getfile.aspx?itemid=56f6368e-54f0-11e7-941a-005056bc530c>>.

⁹Et eksempel fra „starfaskrá“ (occupational titles).

¹⁰Betænkning om ligebehandling af sprog i en digital tidsalder: http://www.europarl.europa.eu/doceo/document/A-8-2018-0228_DA.pdf.

- Birna Arnbjörnsdóttir og Hafdís Ingvarsdóttir. (2018). *Language Development across the Life Span. The impact of English on Education and Work in Iceland*. Cham: Springer.
- Dóra Hafsteinsdóttir. (1998). Fagorðaförðinn. *Orð og tunga*. Reykjavík: Orðabók Háskólans. S. 45–52.
- Dóra Hafsteinsdóttir og Sigríður Harðardóttir. (2001). Íslenska alfræðiorðabókin. Lýsing á viðamiklu orðabókarverkefni. *Orð og tunga*. Reykjavík: Orðabók Háskólans. S. 13–22.
- Eiríkur Rögnvaldsson, Kristín M. Jóhannsdóttir, Sigrún Helgadóttir, Steinþór Steingrímsson, & META Net. (2012). *Íslensk Tunga á Stafrænni öld = The Icelandic Language in the Digital Age*. Georg Rehm, Hans Uczkoreit (ed.). Heidelberg: Springer Verlag. <<http://www.meta-net.eu/whitepapers/e-book/icelandic.pdf>>
- Halldór Halldórsson og Baldur Jónsson. (1993). Íslensk málnefnd 1964–1989. *Rit Íslenskrar málnefndar* 8. Íslensk málnefnd: Reykjavík.
- Mennta- og menningarmálaráðuneytið. (2009). *Íslenska til alls. Tillögur Íslenskrar málnefndar að íslenskri málstefnu samþykktar á Alþingi 12. mars 2009*. <http://islenskan.is/images/Islenka_til_alls.pdf>
- Sigurður Jónsson. (2007). *Det vilda tänkandet och det kultiverande. Isländsk fackspråklig språkvård med tyndpunkt på första hälften av 1900-talet*. Acta Wasaensia, Nr. 184, Språkvetenskap 34, Universitas Wasaensis. Vaasa.
- Sigrún Helgadóttir og Kristján Árnason. (1991). *Áætlun um tilhögun iðorðastarfsemi. Tillögur*. Reykjavík: Íslensk málnefnd, Menntamálaráðuneytið.
- Markáætlun í tungu og tækni <<https://www.rannis.is/sjodir/rannsoknir/markaaetlun-i-tungu-og-taekni>> (<<https://en.rannis.is/funding/research/the-strategic-research-and-development-programme-for-language-technology/>>).
- Máltæknisjóður (Icelandic Language Technology Fund) <<https://www.rannis.is/sjodir/rannsoknir/maltaeknisjodur>> (<https://en.rannis.is/funding/research/icelandic-language-technology-fund/>) (30. september 2019).
- Machine translation for public administrations - eTranslation https://ec.europa.eu/info/resources-partners/machine-translation-public-administrations-etranlation_en (30. september 2019).
- Reglugerð um Íslenska málnefnd og starfsemi Íslenskrar málstöðvar, nr. 159/1987. [Bekendtgørelse om Islandske sprognavn og Islandske sprogsekretariats virksomhed]. <https://www.reglugerd.is/reglugerdir/allar/nr/159-1987>.
- Reglur um um starfsemi Íslenskrar málnefndar nr. 49/1965. [Bekendtgørelse om Islandske sprognavns virksomhed.] [https://brunnur.stjr.is/mrn/logogregl.nsf/b94f17dde835b9a1002568e70043bced/7c537b15201b48ae002568690032261e/\\$FILE/R491965.pdf](https://brunnur.stjr.is/mrn/logogregl.nsf/b94f17dde835b9a1002568e70043bced/7c537b15201b48ae002568690032261e/$FILE/R491965.pdf).

7. Andre kilder

- Betækning om ligebehandling af sprog i en digital tidsalder, <http://www.europarl.europa.eu/doceo/document/A-8-2018-0228_DA.html#title1>. (30. september 2019).
- eTranslation TermBank: <<https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/2016-eu-ia-0122>> (30. september 2019).
- Íðorðabankinn. <idord.is> (30. september 2019).
- Íðorðabankinn, informationsside om arkitektur: <<https://idord.arnastofnun.is/leit/ordabok/ARKITEKT>> (30. september 2019).
- Íðorðabankinn, informationsside om meteorologi: <<https://idord.arnastofnun.is/leit/ordabok/VO>> (30. september 2019).
- Lög um endurnot opinberra upplýsinga, nr. 45/2018 [lov om videreanvendelse af offentlige informationer], <<https://www.althingi.is/lagas/nuna/2018045.html>>.
- Lög um Íslenska málnefnd, nr. 80/1984 [lov om Islandske sprognavn].
- Lög um Íslenska málnefnd, nr. 2/1990 [lov om Islandske sprognavn], <<https://www.althingi.is/lagas/132a/1990002.html>>.
- Lög um Stofnun Árna Magnússonar í íslenskum fræðum, nr. 40/2006, [lov om Árni Magnússon-instituttet for islandske studier]. <<https://www.althingi.is/lagas/nuna/2006040.html>>.

Høsting av digitaliseringens frukter: Hvordan begrepskataloger kan bidra til språkteknologi

Marit Torseth, Jim Yang, Ole Våge

seniorrådgiver, seniorrådgiver, seniorrådgiver

Brønnøysundregistrene, Direktoratet for forvaltning og ikt (Difi), Språkrådet

Brønnøysundregistrene, Postboks 900, 8910 Brønnøysund; Difi, Postboks 1382 Vika, 0114 Oslo; Språkrådet, Postboks 1573 Vika, 0118 Oslo

marit.torseth@brreg.no, jim.yang@difi.no, ole.vage@sprakradet.no

Offentlig sektor i Norge har satt i gang et omfattende arbeid med å definere og beskrive begreper i forbindelse med digitalisering og deling av data i maskinlesbar form. Dette arbeidet kan gjøre det mulig å levere store mengder ett- og flerspråklige termressurser til språkteknologisk gjenbruk.

“Kun én gang” er et av prinsippene som ble fastslått i stortingsmeldingen *Digital agenda for Norge — IKT for en enklere hverdag og økt produktivitet* (2016). Det innebærer at offentlig sektor i mye større grad enn i dag skal gjenbruke data som allerede finnes. For blant annet å operasjonalisere “kun én gang” og gjenbruke data pågår det en stor satsing på informasjonsforvaltning i offentlig sektor. Derfor er det etablert en felles datakatalog med beskrivelser av datasett, begreper, API-er og informasjonsmodeller.

En av de viktigste forutsetningene for å gjenbruke data er at data er godt beskrevet, inkludert begrepene som dataene representerer. Om dataene skal gjenbrukes på tvers av virksomheter, må beskrivelsene av begrepene dessuten være standardiserte. Det er derfor fastsatt flere standarder

som beskriver hva som skal og kan tas med i begrepsbeskrivelser, samt hvordan begrepsbeskrivelser kan gjøres maskinelt tilgjengelig. Et viktig poeng er at standardene også legger til rette for flerspråklige beskrivelser av begreper.

Standardene utgjør en grunnmur som har gjort det mulig å utvikle en felles termportal for offentlig sektor, *Felles begrepskatalog*, som inneholder begreper fra flere store virksomheter. Felles begrepskatalog gjør det mulig å se likheter og forskjeller i begrepsdefinisjoner mellom ulike virksomheter, noe som er nødvendig når de samhandler med hverandre. Felles begrepskatalog gir også tilgang til et stort antall begreper fra offentlig forvaltning og gjør det mulig å høste dem til videre gjenbruk i språkteknologiske ressurser, f.eks. maskinoversettelse mellom norsk og engelsk og mellom bokmål og nynorsk. Gjennom Felles begrepskatalog vil språkbanker få tilgang til oppdaterte termressurser i sanntid.

NordTerm 2019



Landerappporter

Beretning fra Terminologigruppen i Danmark 2017-19

Hanne Erdman Thomsen

Terminologigruppen i Danmark

Introduktion

Terminologigruppen i Danmark er en paraplyorganisation med repræsentanter for de vigtigste danske terminologi-interessenter i og uden for Danmark, og det er Terminologigruppen som repræsenterer Danmark i NORDTERM. Der er mere information om Terminologigruppen på hjemmesiden:

<http://blog.cbs.dk/terminologigruppen/>

Denne rapport omhandler de terminologirelevante aktiviteter som gruppens medlemmer er involveret i, eller som vi har kendskab til. Indledningsvis giver rapporten et overordnet billede af hvordan situationen er for terminologi Danmark.

Fremmedsprogsuddannelser i Danmark

Siden rapporten i 2017 har der været et lille lyspunkt for fremmedsprogsuddannelser i Danmark, idet der er udarbejdet en national strategi for styrkelse af fremmedsprog i uddannelsessystemet (<https://ufm.dk/publikationer/2017/strategi-for-styrkelse-af-fremmedsprog-i-uddannelsessystemet>), og der er etableret et nationalt center for fremmedsprog, NCFF (<https://ncff.dk/>), med sæde i Århus og København. NCFF har til formål at understøtte og styrke fremmedsprog på alle niveauer i uddannelsessystemet, således at flere elever og studerende i fremtiden vil vælge fremmedsprog og opnå gode fremmedsprogskompetencer. Centrets hovedfokus er foreløbig at fordele ressourcer til projekter vedrørende fremmedsprog i folkeskolen og gymnasiet samt overgangen fra gymnasiet til sproguddannelserne og fremmedsprogsundervisning generelt. Der er dog fortsat tilbagegang i søgningen til fremmedsprogsuddannelserne.

Med hensyn til undervisning i terminologi er situationen den at der nu kun er kurser på Aarhus Universitet og Syddansk Universitet, det sidste kursus på CBS Handelshøjskolen i København sluttede i juni 2019. DANTERM Research afholder kurser og workshops om begrebsmodellering i samarbejde med eksterne rekvirenter.

På vej mod en dansk termbank?

I beretningsperioden 2015-17 var der en høring i Folketingets kulturudvalg om sprogteknologi, hvor en dansk termbank også blev bragt på banen. I denne beretningsperiode har kulturministeren nedsat et Sprogteknologiudvalg under Dansk Sprognævn (<http://sprogtek2018.dk/>), som skulle kortlægge behovet og mulighederne for at bruge dansk i teknologier baseret på kunstig intelligens og afklare perspektiverne for en dansk termbank. Udvalget har udarbejdet en rapport, Dansk Sprogteknologi i verdensklasse.

Terminologigruppen i Danmark bidrog til arbejdet med gennemførelsen af en spørgeskemaundersøgelse om terminologi og behovet for en national termbank. Der blev blandt andet spurgt hvilke informationer en national termbank bør rumme, og her blev der blandt andet udtrykt ønske om betydningsbeskrivelser og begrebssystemer, taksonomier og tesauri.

Rapporten anbefaler at der oprettes en dansk sprogbank, herunder en dansk termbank. Rapporten kom i april 2019, og samtidig besluttede den daværende regering at der skulle afsættes knap 30 mio. kr. over de kommende fire år til at lave en fælles dansk sprogresurse, som skal være frit tilgængelig for virksomheder, offentlige myndigheder og forskere. Siden da har der været regeringsskifte, så i skrivende stund er det uvist om bevillingen vil blive en del af den kommende finanslov. Man kan læse mere om rapporten i Sabine Kirchmeiers indlæg på NORDTERM-konferencen 2019.

Aktiviteter hos Terminologigruppens medlemmer

Nedenfor listes de vigtigste projekter og aktiviteter hos en del af Terminologigruppens medlemmer (i alfabetisk rækkefølge). Publikationer er ikke medtaget her, der henvises til medlemsorganisationernes hjemmesider.

Aarhus Universitet, Center for Leksikografi

Regnskabsordbøgerne – (Sandro Nielsen, Lise Mourier, Henning Bergeholtz) løbende revision og udvidelse: Regnskabsordbøgerne er en samling af specialordbøger inden for fagområderne eksternt regnskab, revision, økonomistyring og skatteregnskab. De bygger primært på de internationale regnskabsstandarder (IAS/IFRS), danske, britiske og amerikanske love og vejledninger, samt lærebøger i økonomistyring. De giver hjælp til at forstå, skrive og oversætte regnskabstekster, og de henvender sig til revisorer, regnskabsmedarbejdere, ledere samt medarbejdere i økonomifunktionen i offentlige og private virksomheder, translatører, oversættere, sprogrevisorer, undervisere, økonomi- og sprogstuderende.

<https://www2.ordbogen.com/>

Engelsk-dansk juridisk ordbog – (Sandro Nielsen) Ny online udgave 2019: Ordbogen dækker over de termer og begreber, der er gældende i landområderne England og Wales inklusive et betydeligt antal EU-retlige termer og begreber. Ordbogen behandler termer fra delområderne domstolssystemet, procesretten, arveretten, erstatningsretten, immaterialretten, konkurs- og insolvensretten, kontraktretten, købsretten, "trust"-retten, ejendomsretten, personretten, entrepriseretten, konkurrenceretten, skatteretten, familieretten, mellemmandsretten, ansættelsesretten og strafferetten. Ordbogen indeholder ca. 4.700 termer og definitioner + ca. 4.700 ordforbindelser og deres oversættelser.

<https://www2.ordbogen.com/>

Erhvervs- og Ejendomsordbogen fransk-dansk-fransk: (Liselotte Kruse red.) Ny ordbog baseret på en revision og udvidelse af ejendomsordbogen. Ordbogen henvender sig til medarbejdere i internationale virksomheder (sprogmedarbejdere, salgsmedarbejdere, økonomer, jurister, revisorer, ejendomsmæglere), boligkøbere og -sælgere samt studerende ved universiteter, handelsgymnasier, erhvervsakademier o.l., der har brug for specialiseret terminologi og viden inden for fransk og dansk erhvervskommunikation. Ca. 3000 artikler. <https://www2.ordbogen.com/>

CBS Handelshøjskolen i København, DANTERM Research

Sprogteknologiudvalget. DANTERM Research deltog i Sprogteknologiudvalget, hvor Bodil Nistrup Madsen var udpeget til at varetage arbejdet med at afklare perspektiverne for en dansk termbank, bl.a. ved at stå for en workshop om en national termbank sammen med medarbejdere fra Dansk Sprognævn.

Dansk termbank – en spørgeskemaundersøgelse. DANTERM Research ledede gennemførelsen af spørgeskemaundersøgelsen i forbindelse med udarbejdelsen af Sprogteknologiudvalgets rapport.

eTranslation Termbank. DANTERM Research har bidraget til indsamlingen af terminologi til projektet, hvis danske del blev gennemført af Københavns Universitet, Center for Sprogteknologi.

Dansk Sprognævn

Sprognævnet har lagt et stort arbejde i Sprogteknologiudvalgets rapport, som er omtalt ovenfor.

Derudover bliver der arbejdet på et projekt om **grammatiske termer**. En arbejdsgruppe bestående af Lars Heltoft, Eva Skafte Jensen, Ida Elisabeth Mørch og Jørgen Schack arbejder på at lave et opslagsværk for grammatiske termer. Det skal være en afløser for Grammatisk talt (GT) fra 1996. Det nye opslagsværk vil blive noget længere end GT, og det vil være koncentreret om grammatiske termer. (Den nuværende GT indeholder en del stof om bogstaverne, om tegnsætning ved replikker og om kommatering; dette vil udgå). Opslag i Grammatisk talt bliver set igennem og nyredigeret hvor nødvendigt. Desuden vil der blive føjet en del nye termer til, især fra den grammatikrelevante pragmatik (fx 'sproghandling' og det tekstpragmatiske begreb 'fokus'). Det er hensigten af værket stadig skal være et opslagsværk med forklaringer og eksempler, ikke en lærebog. Når vi vurderer at det kan være nyttigt, vil der dog være ekstra forklaringer af sprog- og traditionshistorisk eller af praktisk karakter (fx hvorfor genitiv ikke er en kasus; hvordan man afgør om noget er en apposition; og lign.). Det overvejes også at føje forslag til videre læsning til. Målgruppen er elever på ungdomsuddannelser, lærere i folkeskolen og danskstuderende på universiteterne. Planen er at udgive værket som e-publikation såvel som trykt bog; der sigtes efter en udgivelse i slutningen af 2019. På længere sigt vil det også være relevant at knytte online ekstramateriale i form af øvelser til værket.

Europa-Parlamentet, Den danske oversættelsesenhed

En stor begivenhed på termområdet siden juni 2017 må siges at være lanceringen af den nye IATE i december 2018. Generelt viger termarbejdet til fordel for oversættelsesarbejdet, når der er en høj arbejdsbyrde, f.eks. i sommeren 2018 i Kommissionen og i Europa-Parlamentet og Rådet i efterår 2018/forår 2019. Det interinstitutionelle termsamarbejde fortsætter med 3 årlige interinstitutionelle møder og en årlig workshop. Vi var sidst vært for workshoppen i april 2019.

Den stigende brug af maskinoversættelse i EU-institutionerne understreger behovet for integrering af terminologi i oversættelsesprocessen, hvilket især Europaparlamentets centrale termkoordination, TermCoord, er fortalende for.

FORVIR

Netværket FORVIR, Forum for Videnmodellering i Offentligt Regi, mødes fortsat ca. 4 gange årligt. Her præsenteres modelleringsarbejdet hos medlemmerne og fælles tiltag drøftes. FORVIR's hjemmeside findes her: <http://blog.cbs.dk/forvir/>

Københavns Universitet, Center for Sprogteknologi (CST)

eTranslation TermBank. eTranslation TermBank er et EU-projekt som har til formål at identificere, indsamle og tilgængeliggøre terminologiske resurser fra offentlige myndigheder og styrelser i de europæiske lande. Terminologiske resurser vil blive indsamlet inden for domænerne sundhed, erhvervsjura og forbrugerbeskyttelse. Data skal bruges til at udvikle automatiske oversættelsessystemer af høj kvalitet for EU-sprogene på [CEF eTranslation](#)-platformen, som er tilgængelig for offentlige myndigheder. CST er arbejdsprojektleder for den opgave som vedrører indsamling af terminologiske sundhedsdata i de europæiske lande.

NORDTERM 2019. CST har været vært for NORDTERM 2019 konferencen 11.-14. juni 2019.

Syddansk Universitet

FOKUS – FOrum for **K**ompetence**U**dvikling i **S**prog)

På Syddansk Universitet blev sprogprojektet FOKUS oprettet ultimo maj 2018. FOKUS er en enhed, der som kerneområde arbejder med sprogene tysk, engelsk, dansk og kinesisk. Et væsentligt formål med projektet er at tilføre fremmedsprogsundervisningen nye metodiske tilgange, her spiller tilegnelsen af ordforråd en væsentlig rolle. FOKUS er tillige rettet mod opbygningen af term- og vidensbaser. I 2018-2019 har projektgruppen gennemført en kvantitativ og kvalitativ behovsafdækning.

Dansk termbank – en spørgeskemaundersøgelse

Syddansk Universitet bidrog væsentligt til udarbejdelsen af spørgeskemaet om terminologi i forbindelse med rapporten om dansk sprogteknologi og en dansk termbank.

Translatørforeningen og Sprogambassadørerne

En del af Sprogambassadørernes og Translatørforeningens medlemmer udfører terminologiarbejde i forbindelse med deres arbejde enten som selvstændige eller i de virksomheder hvor de er ansat.

Konklusion

Fremmedsprog og især terminologi er stadig trængt på de danske universiteter. Der udføres dog terminologiarbejde i en del virksomheder og offentlige organisationer, og der er fortsat interesse for brug af terminologiske metoder i forbindelse med it-udvikling. Udgivelsen af rapporten om dansk sprogteknologi og den tidligere regerings udmelding om en bevilling til en sprogbank giver håb om at der vil blive taget initiativ til etablering af en dansk termbank.

Rapport från Finland

Terminologicalentralen TSK

Vid TSK finns 8 anställda (7 terminologer och 1 IT-planerare).

För 2019 har TSK fått stöd från undervisnings- och kulturministeriet för allmännyttiga tjänster, såsom termbanken, handbiblioteket, informationsverksamhet (samt webbtidskrift Terminfo) och samarbete.

TEPA är Terminologicalentralens offentliga termbank som är avgiftsfritt tillgänglig på Terminologicalentralens webbplats. Under 2018 gjordes cirka 850 000 sökningar i databasen vilket var över 130 % mer än under 2017. I slutet av 2018 innehöll TEPA över 355 000 termposter. TEPA:s samlingar kompletterades med nya eller uppdaterade ordlistor som har skapats i samarbete med Terminologicalentralen. Dessutom uppdaterades materialet från Vetenskapstermbanken i Finland och EU:s interinstitutionella termdatabas IATE.

TSKs projekt under 2019

Begreppsarbete för digitalisering av statsunderstödsverksamhet

Begreppsarbete för digitalisering av statsunderstödsverksamhet sattes igång på uppdrag av CSC – IT Center for Science Ltd på våren 2018. Begreppsarbete ska stödja utveckling och digitalisering av statsunderstödsverksamheten, vars förberedelser undervisnings- och kulturministeriet och finansministeriet har satt igång med. Termer och definitioner i ordlistan ges på finska. Arbetet fortsätter under 2019.

FPA-termer

Terminologiprojektet med Folkpensionsanstalten (FPA) påbörjades under 2007. Under projektets gång har det publicerats fem upplagor av ordlistan FPA-termer, den senaste femte upplagan i 2018. Termerna och de övriga uppgifterna om begreppen ges på både finska och svenska. Uppdateringsarbetet fortsätter under 2019.

Ontologi för den offentliga sektorns tjänster (JUPO)

Uppdateringen av Ontologi för den offentliga sektorns tjänster (JUPO) sattes igång på våren 2019. Den uppdaterade versionen ska publiceras i Nationalbibliotekets Finto-service. Termer i ontologin ges på finska.

Ontologi för museibranschen och Konstindustriella ontologin

I början av 2019 påbörjade Museiverket och Terminologicalentralen ett gemensamt projekt för att uppdatera och utöka MAO/TAO, ontologi för museibranschen och konstindustriella ontologin, med begrepp som gäller arkeologiska föremål. Termer i ontologin ges på finska. Arbetet ska bli färdigt på hösten 2019.

Ordlista för digitalt stöd för utbildningsanordnare

Projektet Bästa digistöd (Parasta digitukea) anlätade Terminologicalentralen under 2018 för terminologiarbete som rör utbildningsavtal och läroavtal inom yrkesutbildning. Ordlistan ska stödja interoperabilitet mellan informationssystem som används för administration av utbildningsavtal och läroavtal. Termer och definitioner ges på finska och ekvivalenter på svenska. Arbetet fortsätter under 2019.

Ordlista för informationshantering inom socialvården

Den första versionen av Sosiaalialan tiedonhallinnan sanasto (Ordlista för informationshantering inom socialvården) publicerades i 2015. Under 2018 anlätade Institutet för hälsa och välfärd Terminologicalentralen att komplettera och uppdatera ordlistan med begrepp som rör till exempel moderskapsutredning. I mars 2019 blev version 5.0 av ordlistan färdig. Termer och definitioner i ordlistan ges på finska.

Ordlista för projektet Tillstånd och tillsyn

Målet med projektet Tillstånd och tillsyn är att modernisera verksamhetssätten för tillstånds- och tillsynsverksamheten genom att utnyttja digitalisering. Den första versionen av ordlistan för projektet blev färdig vid slutet av 2017. I början av 2019 anlätade arbets- och näringsministeriet Terminologicalentralen att komplettera ordlistan med nya begrepp. Ordlistan ska förenhetliga och klargöra begrepp som rör tjänsteverksamhet och tillsyn för att stödja den semantiska interoperabiliteten hos informationssystem. Termer och definitioner ges på finska.

Ordlista för Skatteförvaltningen

På våren 2018 anlätade Skatteförvaltningen Terminologicalentralen för att utarbeta en ordlista på centrala begrepp som rör personbeskattning och som Skatteförvaltningens kunder mest behöver. Den första versionen av ordlistan

blev färdig vid början av 2019 och arbetet med den andra versionen har påbörjats. Ordlistan innehåller termer och definitioner på finska.

Ordlista för social- och hälsovården och arbets- och näringsstjänsterna

På våren 2019 påbörjade Terminologicentralen med arbets- och näringsministeriet ett projekt för att definiera begreppen som är gemensamma för social- och hälsovården och arbets- och näringsstjänsterna. Ordlistan ska innehålla termer och definitioner på finska.

Ordlista för Tullen

Terminologicentralen konsulterar terminologiarbete som påbörjades på den finska tullen under 2018. Tullen har som syfte att reda ut sina begrepp och harmonisera de begrepp som är gemensamma till den offentliga förvaltningen. Ordlistan ska innehålla termer och definitioner på finska. Arbetet fortsätter under 2019.

Ordlista om jäv

På våren 2019 anlidade Kommunförbundet Terminologicentralen med terminologiarbete som rör jävet. Ordlistan ska stödja byggandet av en chatbot som svarar på enkla frågor om jävet. Ordlistan ska innehålla termer och definitioner på finska.

Ordlista om övergripande säkerhet

Ordlista om övergripande säkerhet publicerades för första gång i 2014, och den uppdaterades under 2017. I februari 2019 anlidade Säkerhetskommittén Terminologicentralen för att påbörja ett projekt vars syfte är att uppdatera ordlistans begreppsbeskrivningar och utöka ordlistan med nya säkerhetsbegrepp. De nya begreppen i ordlistan ges termer och begreppsbeskrivningar på finska och termekvivalenter på svenska och engelska. Som en del av uppdateringsprojektet ska Ordlista om övergripande säkerhet förenas med Ordlista om cybersäkerhet (TSK 52), som publicerades i 2018.

Ordlista över centrala begrepp inom fastighets- och byggnadsbranschen

Miljöministeriet har anlitat Terminologicentralen för ett projekt som påbörjades på våren 2018. Projektets mål är att uppdatera och utöka ordlistan över centrala begrepp inom fastighets- och byggnadsbranschen som först publicerades hösten 2016. Termer i ordlistan ges på finska, svenska och engelska och definitioner på finska. Den nya versionen av ordlistan planeras vara färdig på hösten 2019.

Tekla Structures glossary

Den första versionen av den engelska ordlistan Tekla Structures glossary, som uppgjordes på uppdrag av Trimble Solutions Oy (f.d. Tekla Oyj), blev färdig i slutet av 2008 och har uppdaterats sedan dess vid behov. I 2018 fick Terminologicentralen igen ett uppdrag att uppdatera ordlistan, som innehåller engelska termrekommendationer och definitioner, i samarbete med experterna på Trimble Solutions. Arbetet fortsätter i 2019.

Terminologiarbete för Trafikverket

Trafikverket anlidade Terminologicentralen år 2018 för terminologiarbete med två olika mål: att stödja projektet Velho som har som syfte att utveckla trafikledsnätverk, och att främja strategiplanering av det nya Trafikledsverket som ska börja sin verksamhet i början av 2019. Arbetet blev färdigt i februari 2019.

Tietotekniikan termitalkoot – den finska datatermgruppen

Datatermgruppen påbörjade sin verksamhet under 1999 och fortsätter sitt arbete även i år. Arbetet finansieras med medel från undervisnings- och kulturministeriet. Finska och engelska termer publiceras på Terminologicentralens webbplats.

Undervisnings- och utbildningsordlista

Undervisnings- och kulturministeriets terminologiarbete påbörjades år 2010. Arbetets syfte är att stödja olika informationsförvaltningsprojekt som koordineras av ministeriet. Den första versionen av Undervisnings- och utbildningsordlista publicerades på våren 2018. Till största delen ges termer och definitioner i ordlistan på finska, men en del av termerna ges ekvivalenter även på svenska och engelska. Arbetet fortsätter under 2019.

Projekt som blev färdiga under 2018

Ontologiprojekt för Trafikverket

Trafikverket och Terminologicentralen TSK påbörjade år 2016 ett gemensamt ontologiprojekt som hade som syfte att utöka Merenkulkualan ontologia (MERO; ontologin för sjöfart) med begrepp som gäller vägtrafik. Resultatet blev Liikenteen ontologia (LIIKO; ontologin för trafik) som publicerades i Nationalbibliotekets Finto-service hösten 2018. Ontologin har cirka 2 700 begrepp med termer på finska.

Ordlista för Patent- och registerstyrelsen

Terminologicentralen och Patent- och registerstyrelsen (PRS) påbörjade gemensamt terminologiarbete år 2011. Syftet med arbetet var att förenhetliga användning av begrepp inom PRS samt främja den semantiska interoperabiliteten av datasystem mellan PRS och dess samarbetspartner. Ordlistan Patentti- ja rekisterihallituksen sanasto, som blev färdig vid slutet av 2018, innehåller termer och definitioner på finska och termekvivalenter på svenska och engelska.

Ordlista om cybersäkerhet

Sekretariatet för säkerhetskommittén och Försörjningsberedskapscentralen påbörjade våren 2017 ett terminologiprojekt tillsammans med Terminologicentralen. Under projektets gång utarbetades en ordlista som omfattar cirka 70 centrala begrepp som anknyter till cyber- och informationssäkerhet. Ordlista om cybersäkerhet publicerades sommaren 2018 och innehåller definitioner på finska samt termer på finska, svenska och engelska.

Ordlista om likabehandling i social- och hälsojänster

Institutet för hälsa och välfärd (THL) anlätade Terminologicentralen på våren 2018 för att utarbeta en ordlista om centrala begrepp som rör likabehandling i social- och hälsojänster på finska. Dessa begrepp behövs i utvecklandet av utvärdering av social- och hälsojänster. Ordlistan blev färdig sommaren 2018.

Ordlistor för Statistikcentralen

Under 2018 utarbetade Terminologicentralen två ordlistor för Statistikcentralen. Ordlistan Työhön liittyviä tilastokäsitteitä omfattar centrala begrepp som gäller arbetsliv och personens ekonomi medan ordlistan Energiaan ja päästöihin liittyviä tilastokäsitteitä innehåller begrepp som gäller energi och utsläpp. Båda ordlistorna har termer och definitioner på finska.

Termer och ontologier för geografisk information

Projektet Termer och ontologier för geografisk information sattes igång på uppdrag av Lantmäteriverket i 2010. Projektets mål är att klargöra begrepp och rekommendera termer som anknyter sig till INSPIRE-direktivet (direktivet om infrastruktur för rumslig information i Europeiska gemenskapen) samt främja både utveckling av den nationella infrastrukturen för geografisk information och genomförandet av den nationella strategin för geografisk information. Den fjärde upplagan av Ordlista för geoinformatik publicerades våren 2018. Ordlistan innehåller definitioner på finska och termerna på finska, svenska och engelska.

Uppdatering av Byggnadsklassificering 1994

Statistikcentralen anlätade Terminologicentralen hösten 2016 att uppdatera Byggnadsklassificering 1994. Både klassificeringens struktur och namn till byggnadsklasserna förändrades och beskrivningar till byggnadsklasserna utarbetades. Den nya klassificeringen, Byggnadsklassificering 2018, publicerades i december 2018 på finska, svenska och engelska.

Övriga projekt

Nationalbiblioteket har anlitat Terminologicentralen att utveckla ett verktyg som stödjar uppdatering av ontologier. Utvecklingsarbetet, som påbörjades i 2018, genomförs i samarbete med Nationalbibliotekets Finto-projekt.

Den terminologiska verksamheten vid översättnings- och språksektorn vid statsrådets kansli

Aktuellt 2018-2019

Statsrådets kansli publicerade följande handböcker och rekommendationer:

- *Svenskt lagspråk i Finland - ordlistan* publicerades i statsrådets termbank Valter och som öppna data i formatet TermBase eXchange. Ordlistan innehåller cirka 1590 centrala ord, uttryck och fraser som förekommer i finska lagtexter samt rekommendationer till deras motsvarigheter på svenska.
www.valter.fi
<https://www.avoindata.fi/data/sv/dataset/ruotsinkielinen-lakikieli-slaf-sanasto>
- Ordlistor med terminologin för aktuella lagstiftningsprojekt: social- och hälsovårdsreformen, de nya landskapen och underrättelseverksamheten
Centrala termer i social- och hälsovårdsreformen fi-sv-en:
<https://vnk.fi/documents/10616/3457861/Keskeisiä+sote-uudistuksen+termejä+fi-sv-en>
Termer med anknytning till förvaltningsstrukturen i Finland och till landskapsreformen samt landskaps- och kommunnamn fi-sv-en-(ru):
[https://vnk.fi/documents/10616/3457861/Suomen+hallintorakenteeseen+ja+maakuntauudistukseen+liittyviä+termejä+sekä+maakuntien+ja+kuntien+nimet+fi-sv-en-\(ru\)](https://vnk.fi/documents/10616/3457861/Suomen+hallintorakenteeseen+ja+maakuntauudistukseen+liittyviä+termejä+sekä+maakuntien+ja+kuntien+nimet+fi-sv-en-(ru))
Ordlista över underrättelsetermer fi-sv-en:
<https://vnk.fi/documents/10616/3457861/Tiedustelutermien+sanasto+fi-sv-en>
- Stilregler för att översätta och skriva på engelska, franska och tyska under Finlands EU-ordförandeskap
Presidency Style Guide:
<https://vnk.fi/documents/10616/12801509/Englannin+kielen+EU2019-tyyliopas.pdf/559c05fb-868f-4bd6-a21b-96bf0d9e1172/Englannin+kielen+EU2019-tyyliopas.pdf.pdf?version=1.0>
Übersetzen von Online-Inhalten für den EU-Ratsvorsitz:
<https://vnk.fi/documents/10616/12801509/Saksan+kielen+EU2019-tyyliopas.pdf/90648412-7a5c-9614-0d71-d47d59e368a4/Saksan+kielen+EU2019-tyyliopas.pdf.pdf?version=1.0>
Traduction de contenus en ligne pour la présidence tournante du Conseil de l'Union européenne:
<https://vnk.fi/documents/10616/12801509/Ranskan+kielen+EU2019-tyyliopas.pdf/da5d5b45-6051-0f03-12e2-5c68315a92a4/Ranskan+kielen+EU2019-tyyliopas.pdf.pdf?version=1.0>

Aktuellt just nu:

- terminologiarbete som stöder Finlands EU-ordförandeskap och den nya regeringens verksamhet särskilt ur översättningens synvinkel
- att testa maskinöversättning

Rapport fra Island

Ágústa Þorbergsdóttir

Árni Magnússon-instituttet for islandske studier
Laugavegi 13, IS-101 Reykjavík
agusta.thorbergisdottir@arnastofnun.is

1. Árni Magnússon-instituttet for islandske studier

Árni Magnússon-instituttet for islandske studier (ÁM-instituttet) er et center for termarbejde i Island. Hos ÁM-instituttet hører terminologien under sprogrøgtsafdelingen. Genstandsområde:

- termbanken, *idord.is*
- publicering af terminologiske samlinger
- vejledning til fagfolk/termudvalg om terminologiske arbejdsmetoder og om sproglige problemer
- rådgivning (svar på spørgsmål fra almenheden og myndighederne om oversættelser af fagtermer)
- praktisk tjeneste for termudvalg, fx mødelokale og arbejdsplads

2. Termbanken *idord.is*

For nylig blev den tredje version af termbanken lanceret. Den er tilgængelig på webadressen *idord.is* og indeholder nu 60 termsamlinger og 170.000 termposter. Termbankens navn er blevet ændret fra *Orðabanki Íslenskrar málstöðvar* (Islandsk sprogsekretariats ordbank) til *Íðorðabankinn* (Termbanken). Navneændringen var tiltrængt, men ved en lovgivning der trådte i kraft i 2006 blev Íslensk málstöð (Islandsk sprogsekretariat) forenet med fire andre institutioner på sprog- og litteraturområdet i Stofnun Árna Magnússonar í íslenskum fræðum (Árni Magnússon-instituttet for islandske studier) som overtog dets opgaver og pligter.

I den nye version er brugergrænsefladen er enklere og søgesiden er forenklet. Fuzzy søgning er nu implementeret hvis brugeren kommer til at indtaste en stavfejl og søgning er ved hjælp af autofuldførelse (autocomplete). En nyhed er at nu kan man vælge en bestemt termsamling og se dens indhold. Man kan desuden vælge at scrolle termsamlingerne efter islandsk eller udenlandsk alfabetisk orden.

Nyt materiale bliver hele tiden tilføjet i ordbanken, fx har termudvalgene inden for medicin, fritidspædagogik, public health og håndarbejde revideret deres termsamlinger og tilføjet nye termposter.

3. Udgivelser

I tidsrummet 2017–2019 blev to termsamlinger for anatomi publiceret: *Orðasafn í líffærafræði, III Æðakerfið* (Anatomisk terminologi, årer) blev publiceret i 2017 og *Orðasafn í líffærafræði, IV. Taugakerfið* (Anatomisk terminologi, nervesystemet) blev publiceret i 2019. De er begge to i serien *Íðorðarit Stofnunar Árna Magnússonar í íslenskum fræðum*. Disse termsamlinger opdeles i en systematisk ordnet del med engelsk-islandske og islandsk-engelske ordlister. Denne opdeling skal opfylde et behov som de som er særligt interesserede i menneskets anatomi

kan have, fx studenter, lærere og ansatte i sundhedssektoren. Formålet med denne inddeling af ordlisten er at klargøre forskellen mellem begreber som ligger tæt på hinanden.

Orðasafn í tómstundafræði (Termsamling i fritidspædagogik) blev publiceret i 2019 med bistand fra Árni Magnússon-instituttet for islandske studier.

Disse tre ovennævnte termsamlinger fås både i trykt udgave og som gratis pdf-udgave.

Termsamlingen *Nöfn háplöntuætta* (Plantenavn) blev publiceret i bogform i 2019. Termerne er revideret i overensstemmelse med ny forskning og ændringer der er foretaget i definitioner af mange plantefamilier. Termsamlingen *Nöfn háplöntuætta* findes også i *Íðorðabankinn*.

4. Termudvalgene

Det er i år 100 år siden det første termudvalg blev oprettet af Ingeniørforeningen. Nu er der registreret omkring 50 termudvalg, hvoraf halvdelen er aktive.

5. Terminologiforeningen

Terminologiforeningen som blev grundlagt i 2010, har som hovedformål at fremme arbejdet inden for terminologi i Island og at repræsentere terminteresserede over for myndigheder og de institutter som ifølge loven har det som deres opgave at fremme arbejdet med terminologi. For at nå foreningens formål arrangerer man blandt andet seminarer og diskussionsmøder for medlemmerne. Disse møder afholdes cirka fire gange om året.

6. eTranslation TermBank

Árni Magnússon-instituttet for islandske studier deltog i projektet eTranslation TermBank som havde til opgave at skaffe terminologiske ressourcer fra offentlige institutioner i hele Europa til at forbedre kvaliteten af EU's automatiske oversættelsessystem, CEF. Terminologiske ressourcer blev indsamlet inden for domænerne e-Justice (retsinformation), eHealth (sundhed) og ODR Open Dispute Resolution (forbrugerbeskyttelse).

7. Undervisning i terminologi

Inden for faget oversættelse ved Islands Universitet er der tilbudt et kursus i terminologi (5 ECTS). Árni Magnússon-instituttet giver desuden termudvalgene vejledning om udarbejdelsen af termsamlingerne.

Statusrapport for Norge

Marianne Aasgaard

Språkrådet

marianne.aasgaard@sprakradet.no

1. Språkrådet

Språkrådet er nasjonalt samordningsorgan for terminologiarbeid i Norge og har tre ansatte som har terminologi og fagspråk som oppgaver. Dette arbeidet ligger under seksjon for fagspråk og språk i samfunn og høyere utdanning.

Språkrådet jobber på strategisk nivå med å legge til rette for terminologiarbeid i flere sektorer. Målet er å øke utvikling og tilgjengeliggjøring av terminologi og fagspråk på norsk, både bokmål og nynorsk.

De strategisk prioriterte områdene for Språkrådet er

- begrepsarbeid i forbindelse med digitalisering av offentlig sektor (felles begrepskatalog)
- terminologiarbeid i helsesektoren
- terminologiprojekter i universitets- og høyskolesektoren
- teknisk terminologi

Språkrådet arbeider også med å gjøre terminologiske ressurser tilgjengelige for gjenbruk i språkteknologiske løsninger.

Språkrådet er også i ferd med å utvikle et e-læringskurs i terminologisk metode, slik at vi kan nå ut til flere med en grunnleggende opplæring i terminologi.

1.1 Begrepsarbeid i offentlig sektor

1.1.1 Felles begrepskatalog

Språkrådet er involvert i arbeidet med felles begrepskatalog i staten. Felles begrepskatalog har utgangspunkt i felles datakatalog som inneholder en oversikt over datasettene i offentlige virksomheter. Felles begrepskatalog inneholder beskrivelser av datasettene. Språkrådet har vært involvert i arbeidet med flere standarder for dette arbeidet :

- standard for begrepsbeskrivelser
- standard for begrepskoordinering og -harmonisering
- standarder for tilgjengeliggjøring av begrepsbeskrivelser (SKOS og TBX)

1.1.2 Terminologiarbeid i helsesektoren

Direktoratet for e-helse har opprettet et eget terminologisenter med to ansatte, og de arbeider blant annet med å oversette kodeverket SNOMED til norsk. Språkrådet var involvert i arbeidet med å utarbeide språklige retningslinjer for dette arbeidet.

1.1.3 Kunnskapsdepartementet (med flere)

Kunnskapsdepartementet har tatt initiativ til og gjennomført et terminologiprojekt for overordnet studieadministrativ terminologi.

1.2 Universitets- og høyskolesektoren

Språkrådet er involvert i flere terminologiprojekter i universitets- og høyskolesektoren. Språkrådet er involvert på flere ulike nivåer, fra å sitte som terminolog i gruppa til kun å støtte prosjektet økonomisk.

Termgruppene i universitets- og høyskolesektoren omfatter flere ulike fagområder:

- kjemi
- celle- og molekylærbiologi
- lingvistikk
- matematikk
- klima
- astronomi
- studieadministrativ terminologi

2. Termportalen

Språkrådet, Universitetet i Bergen og Norges Handelshøyskole (NHH) har blitt enige om en intensjonsavtale for den nasjonale termbasen *Termportalen*. Universitetet i Bergen vil ha ansvaret for driften av termbasen, mens NHH leverer innhold på sine fagområder. Språkrådet vil være koordinator for arbeidet.

Det er også satt i gang et forprosjekt som skal se på hvordan drift og vedlikehold av *Termportalen* skal finansieres.

3. Standard Norge

Språkrådet og Standard Norge har et godt samarbeid. De to virksomhetene inngikk en felles handlingsplan i 2017, og har fulgt opp punktene i denne planen. Hovedutfordringen nå er å få en fast finansieringsordning for oversettelse av standarder og utvikling av terminologi på bokmål og nynorsk til termbasen Snorre.

4. Språklov og stortingsmelding om språkpolitikken

Kulturdepartementet har arbeidet med utkast til lov om språk (språklov), og den er ute på høring høsten 2019. Departementet arbeider parallellt med en stortingsmelding for en ny og oppdatert språkpolitikk i Norge.

Terminologiverksamheten i Sverige

Perioden juni 2017 till juni 2019

Slutet för Terminologicentrum TNC – från flytten in i SIS till nedläggningen

Under sommaren 2017 flyttade TNC:s kansli in hos huvudägaren SIS. Efter att flera medarbetare blivit uppsagda eller valt att sluta återstod då fyra anställda (varav två personer var tjänstlediga för studier på heltid respektive halvtid). Alla administrativa, ekonomiska och datatekniska tjänster köpte TNC från och med då av SIS. Under hösten och vintern valde ytterligare två anställda att säga upp sig. Våren 2018 beslutade SIS att TNC skulle läggas ned, och vid årsskiftet 2018-2019 upphörde verksamheten vid TNC helt.

Från och med budgetåret 2017 hade Språkrådet, en avdelning inom myndigheten Institutet för språk och folkminnen (Isof), fått ett ökat ansvar för det nationella terminologiarbetet, som en del av det övergripande arbetet med svensk språkvård och genomförande av språkpolitiken. Det statliga bidraget för terminologiarbete gick därmed till Språkrådet. Under 2017 överfördes dock medlen helt, och under 2018 delvis, till TNC. När SIS aviserade om nedläggning av TNC kom diskussionerna om hur det ökade terminologiska ansvaret skulle bedrivas av Språkrådet igång på allvar.

För att kunna bygga upp en terminologiverksamhet anställde Språkrådet under hösten 2018 två tidigare TNC-terminologer. Dessutom fick Språkrådet ta över TNC:s bibliotek, som nu finns i Språkrådets lokaler. Vidare förvaltas Rikstermbanken numera av Språkrådet. Och på det statliga arkivet Riksarkivet finns dokumentation från organisationen TNC bevarad för framtiden.

Hur ser terminologi-Sverige ut efter nedläggningen av TNC?

Språkrådet

På Språkrådet finns sedan november 2018 två terminologer, med titeln språkvårdare med inriktning mot fackspråk och terminologi, som alltså har börjat bygga upp terminologiverksamheten vid Språkrådet.

En uppenbar skillnad i verksamhetens inriktning har att göra med att Språkrådet är en myndighet. Det innebär bland annat att man inte kommer att ta större betalda uppdrag, varken av privata aktörer eller av enskilda myndigheter. Språkrådet kommer vidare inte att hålla terminologiutbildningar i den omfattning som TNC gjorde och inte heller ha termrådgivning i någon större utsträckning.

Språkrådets verksamhet kommer alltså i huvudsak rikta sig mot terminologiarbete inom offentlig förvaltning. Terminologerna kommer att stötta det terminologiarbete som finns eller behöver utvecklas på myndigheterna, dock inte leda och medverka i särskilda terminologiutredningar. Meningen är i stället att myndigheterna på egen hand ska ta ansvar för att reda ut sina egna begrepp såsom det föreskrivs i språklagens § 12: "Myndigheter har ansvar för att svensk terminologi inom deras olika fackområden finns tillgänglig, utvecklas och används." Språkrådet har också planer på att få igång myndighetsövergripande terminologiarbete i olika ämnesvisa grupperingar.

Sammanfattningsvis har Språkrådet primärt en samordnande roll och ska därmed arbeta med att främja, stödja och samordna terminologiarbete inom offentlig förvaltning så att språklagens § 12 kan efterlevas.

I detta arbete ingår bland annat följande:

- Språkrådet sprider information, ger råd och utarbetar riktlinjer för terminologiarbete.
- Språkrådet erbjuder introduktionsföreläsningar och inspirationsseminarier för offentlig verksamhet.
- Språkrådet förvaltar och utvecklar Rikstermbanken.
- Språkrådet främjar utbildning och forskning inom terminologiområdet.
- Språkrådet främjar terminologiprojekt inom offentlig förvaltning och samverkar med nationella och internationella terminologigrupperingar.

Övriga terminologi-Sverige

Konsekvenserna av att en organisation med samlad kompetens och ett utbud av olika terminologitjänster gått i graven börjar nu visa sig, liksom nya utmaningar när det offentliga och det privata terminologiarbetet har separerats.

I terminologi-Sverige utanför Språkrådet finns i dagsläget en del terminologer anställda på myndigheter och företag men inte så många terminologikonsulter som tar uppdrag, förutom före detta TNC-terminologer som arbetar som konsulter. Terminologitjänster har börjat säljas och så även terminologutbildningar, men i liten skala. Några terminologigrupperingar finns fortfarande mer eller mindre aktiva: Svenska biotermgruppen, Svenska datatermgruppen och Svenska optiktermgruppen. Men fler terminologiaktörer, nätverk och grupperingar behöver bildas för att en ny fungerande infrastruktur för terminologi-Sverige ska komma till stand, oavsett var terminologiarbetet bedrivs. Potentiella terminologer kan finnas i yrkesföreningar med språkarbetare: Examinerade språkkonsulter i svenska (Ess) och Sveriges Facköversättarförening (SFÖ) m.fl.

Nybildad förening: Terminologiframjandet

Privata men även i viss mån offentliga aktörer som ägnar sig åt terminologiarbete har i dagsläget inte så lätt att hitta professionella terminologer att vända sig till. Den samlade kompetensen som fanns på TNC har splittrats. Mot den här bakgrunden bildade – i december 2018 innan TNC lades ner – ett antal före detta TNC:are föreningen Terminologiframjandet.

Enligt stadgarna ska föreningen verka för en effektiv fackspråklig kommunikation i det svenska samhället, främja utvecklingen och samordningen av olika fackspråk och terminologier, uppmuntra till att de uppmärksammas och studeras. Föreningen ska också framhålla terminologyrket. Medlemmar i föreningen kan vara enskilda personer med intresse för terminologifrågor.

Terminologiframjandet hoppas kunna vara en viktig inspiratör för terminologifrågor i det svenska samhället och en samlande kraft för de personer som arbetar med terminologi (inte bara de som kallar sig ”terminologer”). Läs vidare på föreningens webbplats: www.terminologiframjandet.se

Terminologistandardisering

SIS/TK 115 är aktiv och den svenska versionen av ISO 704 är klar och kommer att komma ut och presenteras vid ett seminarium under 2019. Även ISO 13611 och ISO 10241-1 är översatta och kommer ut inom kort.

Forskning, i grevens tid ...

Deltagare i forskarnätverket ”Termer och terminologisering i svenskt fackspråk” (Göteborgs universitet, Stockholms universitet, Umeå universitet och Vasa universitet) har under perioden bedrivit forskning på TNC:s material. Forskarna har framför allt riktat in sig på de så kallade rundfrågor där olika terminologiska frågor skickades med brev mellan TNC och olika sakkunniga från 40-talet till början av 80-talet. Dessutom har några av forskarna studerat

och spelat in några terminologimöten, för att se hur terminologer under begreppsutredningar samarbetar med deltagarna i olika terminologiska arbetsgrupper.

Språkrådets terminologer ingår, tillsammans med TSK och Institutet för de inhemska språken, i nätverkets referensgrupp.

Fortsatt utbildning i terminologi, ett stort frågetecken

Den typen av öppna kurser i terminologilära som TNC höll i ges inte längre av någon organisation. Tolk- och översättningsinstitutet (TÖI) vid Stockholms universitet, som tidigare haft mycket terminologiutbildning och som TNC samarbetat med, har just nu inte några fristående kurser. Några mindre lärosäten, Linnéuniversitetet och Mälardalens högskola, har dock visat intresse för att hålla terminologikurser.

Så det är en mycket instabil utbildningssituation terminologi-Sverige, både inom högskoleutbildningen och i andra sammanhang ute i samhället.