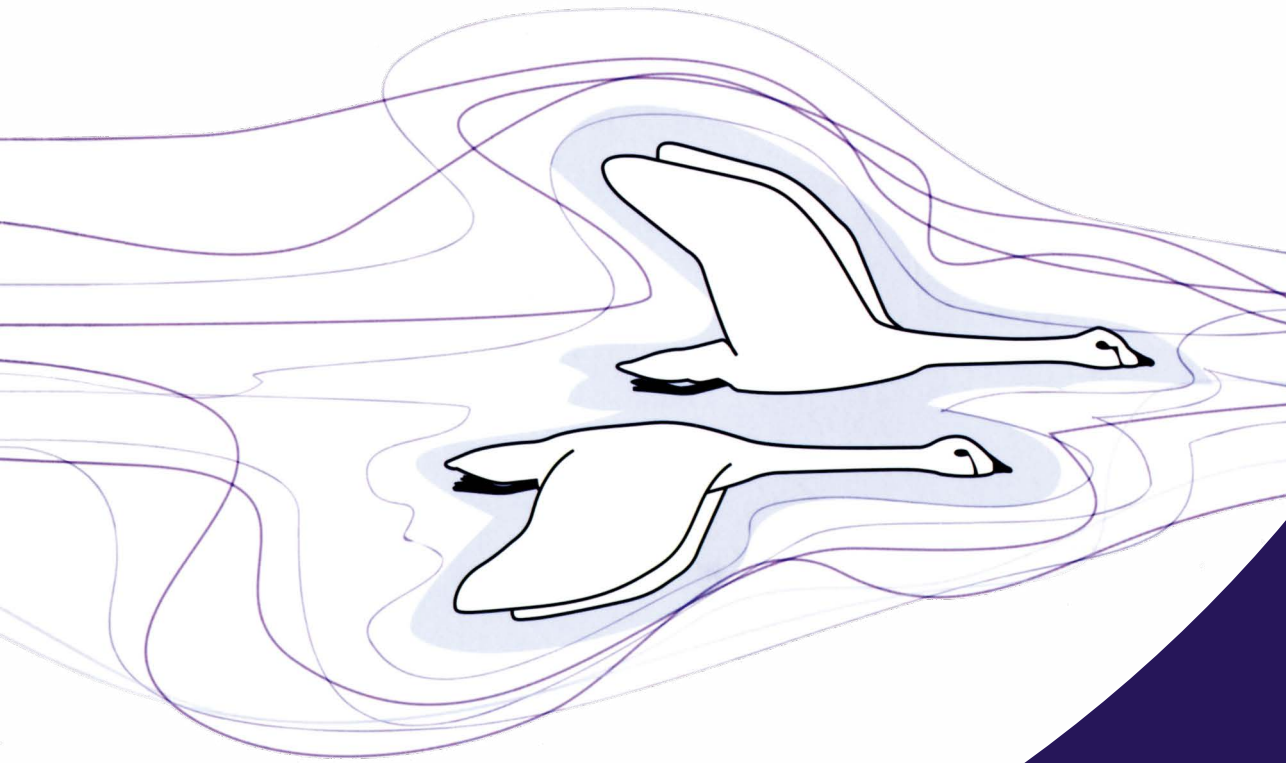


# Språk i Norden 2022

Tema: Språkteknologi





# Språk i Norden 2022

Tema: Språkteknologi



Språk i Norden ges ut  
med stöd av  
Nordiska Ministerrådet



**Språk i Norden ges ut  
med stöd av  
Nordiska Ministerrådet**

Utgiven av Nätverket för språknämnderna i Norden, 2022

Copyright: Författarna och Nätverket för språknämnderna i Norden  
ISSN: 2246-1701

# Innehåll

<b>Förord</b> . . . . .	<b>5</b>
<i>May Wikström</i>	
<b>ELRC, ELG og ELE</b>	
<b>– Europæiske krafttilskud til nordisk sprogteknologi.</b> . . . . .	<b>7</b>
<i>Sabine Kirchmeier</i>	
<b>Robotten</b>	
<b>– Den hundredårige der kom ind på scenen og ikke forsvandt . .</b>	<b>19</b>
<i>Peter Juel Henriksen</i>	
<b>Tillgänglig textning med hjälp av språkteknologi.</b> . . . . .	<b>29</b>
<i>Michael Stormbom</i>	
<b>Donera prat för AI-utveckling till Språkbanken i Finland</b> . . . . .	<b>35</b>
<i>Krister Lindén, Mietta Lennes, Tommi Jaubiainen, Mikko Kurimo, Alekski Rossi</i>	
<b>Språk(teknologi) är nyckeln till (artificiell) intelligens och rättvisa.</b> . . . . .	<b>51</b>
<i>Jörg Tiedemann</i>	
<b>Automatisk målgruppsanpassad textanpassning.</b> . . . . .	<b>67</b>
<i>Evelina Rennes &amp; Arne Jönsson</i>	
<b>Talegenkendelse og sprogteknologi på Færøerne</b> . . . . .	<b>83</b>
<i>Sandra Saxov Lambauge &amp; Iben Nyholm Debess</i>	
<b>Samisk språkteknologi i 2021</b> . . . . .	<b>93</b>
<i>Sjur Nørstebø Moshagen</i>	

<b>Ny språklitteratur 2021</b> . . . . .	<b>103</b>
<i>Danmark, Marianne Rathje</i> . . . . .	103
<i>Finland, Maria Fremer</i> . . . . .	112
<i>Färöarna, Ragnar Sigrunarson</i> . . . . .	120
<i>Grönland, Beatrine Heilmann</i> . . . . .	123
<i>Island, Ágústa Þorbergsdóttir</i> . . . . .	125
<i>Norge, Gunnbild Wiggen</i> . . . . .	129
<i>Sverige, Johanna Ledin</i> . . . . .	137
<b>Adresser till språknämnderna i Norden 2022</b> . . . . .	<b>148</b>

# Förord

*Språk i Norden* är en årsbok utgiven av Nätverket för språknämnderna i Norden.

Publikationen har utkommit under nuvarande namn sedan 1970. Numera publiceras den endast elektroniskt på publikationsportalen Tidsskrift.dk, som tillhandahålls av Det Kongelige Bibliotek i Danmark. Årsboken är öppen för alla och tillgänglig på webbadressen <https://tidsskrift.dk/sin>.

*Språk i Norden* redigeras av de nordiska sekreterarna på språknämnderna (eller motsvarande) i Danmark, Grönland, Finland, Färöarna, Island, Norge och Sverige. Huvudredaktör för årets utgåva är från Finland.

Det nordiska samarbetet är väl etablerat. Sedan 1954 har nätverket för de nordiska språknämnderna vartannat år arrangerat nordiska språkmöten. Dessa samlar bland annat språkvårdare, lexikografer, forskare och centrala aktörer kring ett valt tema.

Som så mycket annat påverkades även mötet år 2021 av pandemin. Det arrangerades därför som ett digitalt evenemang – men även temat hade digitala förtecken och handlade den här gången om språkteknologi.

Artiklarna i *Språk i Norden 2022* bygger på inlägg och föredrag från konferensen. De vittnar om hur teknologin kan bidra till jämlikhet och tillgänglighet men belyser också de specifika utmaningarna och behoven inom språkteknologiskt arbete och forskning.

**Sabine Kirchmeier** redogör i sin artikel för tre europeiska initiativ med stor betydelse för nordisk språkteknologi.

**Peter Juel Henriksens** essä handlar om den teknologiska utveckling som bland annat lett fram till en mänsklig robot som tänker och agerar, åtminstone till synes, som en människa.

**Michael Stormbom** tar fasta på hur information kan tillgängliggöras för grupper med särbehov tack vare språkteknologi som processar tal eller text.

**Krister Lindén, Mietta Lennes, Tommi Jauhiainen, Mikko Kurimo** och **Alexi Rossi** redogör för en kampanj i Finland, där nationella aktörer genom teamarbete samlade in 4 000 timmar vardagsprat för forskning och innovation på bara några månader.

**Jörg Tiedemann** presenterar hur neural maskinöversättning går till och argumenterar för att inte överlåta alla språkteknologiska lösningar till it-jättar och multinationella bolag.

**Evelina Rennes** och **Arne Jönsson** tar fasta på möjligheten att utnyttja språkteknologi för att förenkla texter och mäta hur enkla de är att läsa.

**Sandra Saxov Lamhauge** och **Iben Nyholm Debess** beskriver arbetet med att utveckla taligenkänning och teknologiska språkresurser på Färöarna.

**Sjur Nørstebø Moshagen** synliggör i sin artikel om samisk språkteknologi de specifika och konkreta utmaningar som små språk möts av.

Vi i redaktionen tackar för de insiktsfulla bidragen!

Artikeldelen följs av en översikt över publikationer och ordböcker som getts ut i de nordiska länderna under det gångna året.

Årsboken innehåller dessutom en aktuell adresslista till språknämnderna (eller motsvarande) i Norden.

Redaktionen för Språk i Norden 2022 har bestått av **Ágústa Þorbergsdóttir** (Stofnun Árna Magnússonar í íslenskum fræðum), **Beatrine Heilmann**, (Oqaasileriffik), **Kristin M. Magnussen** och **Ragnar Sigrunarson** (Føroyska málráðið), **Maria Fremer** (Institutet för de inhemska språken, Finland), **Marianne Rathje** (Dansk Sprognævn), **Gunnhild Wiggen** (Språkrådet, Norge) och **Johanna Ledin** (Språkrådet, Sverige).

Helsingfors i april 2022

**May Wikström**

Huvudredaktör



# ELRC, ELG og ELE – Europæiske krafttilskud til nordisk sprogteknologi

*Sabine Kirchmeier*

*I denne artikel præsenterer jeg tre europæiske initiativer med stor betydning for nordisk sprogteknologi. Det drejer sig om European Language Grid (ELG), European Language Resource Coordination (ELRC) og European Language Equality (ELE).*

*Projekterne har det tilfælles at de bygger på et mangeårigt samarbejde mellem de sprogteknologiske forskningsmiljøer og EU-Kommissionen for at skabe god sprogteknologi for alle europæiske sprog.*

## 1. Baggrund

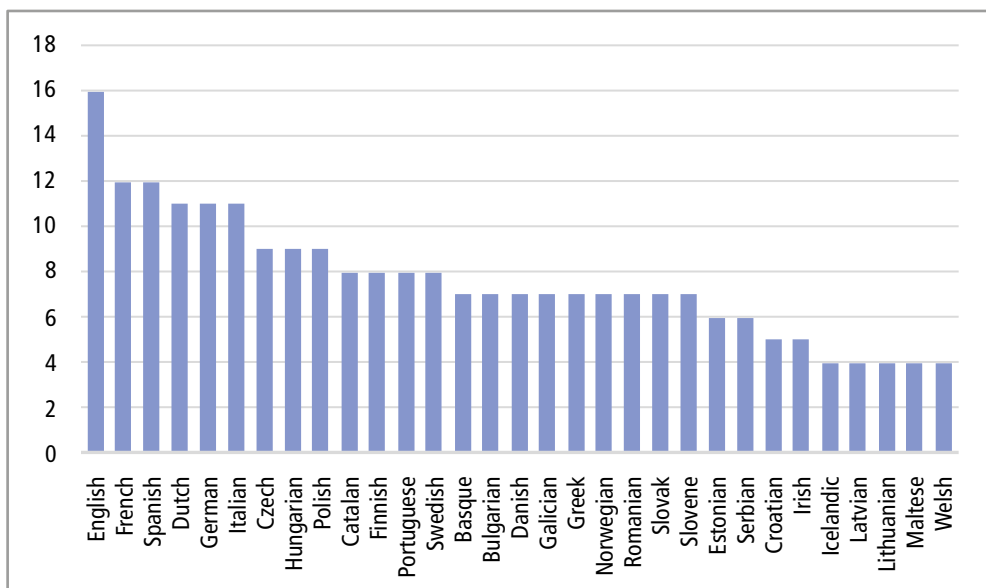
Sprogteknologi er et vanskeligt og komplekst felt – dobbelt komplekst, kan man sige, da det både kræver god indsigt i sprogvidenskab og i datalogi. Det har derfor ofte været en udfordring at forklare fx politikere og borgere, at denne teknologi i modsætning til mange andre teknologiske produkter ikke uden videre fungerer på alle sprog, men kræver en særlig indsats fra hvert enkelt sprogsamfund. Det har gennem årene været svært at få skabt den nødvendige interesse og finansiering for projekterne, og derfor udviklede sprogteknologien for de forskellige sprog sig i helt forskellige tempi alt afhængigt af hvor stor politisk opmærksomhed det lykkedes at skabe.

Bestræbelserne for at indsamle resurser for at skabe god sprogteknologi for de europæiske sprog går efterhånden langt tilbage. I starten drejede de sig om isolerede projekter i de enkelte lande, som førte til store nationale tekstsamlinger. Fx er Språkbanken i Göteborg, som blev oprettet i 1975, et resultat af den type initiativer. Siden kom der europæiske projekter til. Språkbanken var således med i et af de tidlige store fælleseuropæiske sprogresurserprojekter, EU-projektet PAROLE 1996-1997 (EU-Kommissionen 1996), og resultaterne af dette projekt indgår også i Det Danske Sprog- og Litteraturselskabs Korpus. PAROLE involverede 8 europæiske lande, og efterhånden blev der indsamlet resurser på flere og flere sprog dels til forskning og dels til den spirende sprogteknologiske industri.

Det blev hurtigt tydeligt, at der var behov for fælles standarder og fælles

metadata for at de indsamlede tekster og værktøjer kunne fungere sammen. Omkring 2008 kom der gang i det stort anlagte fælles sprogresurseprojekt, CLARIN (<https://www.clarin.eu/>). Her kom for første gang stort set alle europæiske lande med. Clarin var primært anlagt som en forskningsinfrastruktur, og man tænkte som udgangspunkt ikke i kommerciel udnyttelse af de sprogteknologiske data og værktøjer som blev indsamlet.

META-NET (<http://www.meta-net.eu/>) var det første forsøg på at skabe overblik over udviklingen af sprogteknologi i Europa og at arbejde målrettet på at påvirke beslutningstagerne i de europæiske lande og i EU-Parlamentet til at fremme sprogteknologi. META-NET whitepapers (Rehm, G. & Uszkoreit, H. 2012) omfattede således 31 europæiske sprog fra 30 forskellige lande. Rapporten indeholdt en sprogteknologisk status for hvert enkelt sprog og gjorde det muligt at illustrere hvordan de europæiske sprog var understøttet sprogteknologisk, dels i forhold til hinanden og dels – ikke mindst – i forhold til engelsk.



Source: [Key Results](#) and Cross-Language Comparison, META-NET White Paper, data 2012-2014

Figur 1. Oversigt over sprogteknologisk understøttelse for forskellige sprog (2014). Skala: 16 'udmærket understøttelse', 8-11 'moderat eller fragmentarisk understøttelse', 4-6 'svag eller ingen understøttelse'.

Oversigten fra META-NET-rapporten viste tydeligt, hvor lille et sprogteknologisk beredskab der fandtes for de fleste officielle europæiske sprog bortset

fra engelsk på fire centrale områder: tilgængelige tekst- og talesprogsressourcer, automatisk oversættelse, taleteknologi, programmer til tekstanalyse.

Det fremgår af figur 1 at de nordiske sprog er blandt dem der kun er fragmentarisk dækket med finsk og svensk som de bedst forsynede og islandsk med den svageste dækning.

Grafer som denne gjorde indtryk på politikerne, og de blev flittigt brugt nationalt og på europæisk plan for at argumentere for bedre services på de europæiske sprog. Sprogteknologi begyndte langsomt at rykke højere op på dagsordenen, men først med den accelererende udvikling inden for kunstig intelligens kom det egentlige gennembrud.

Selv om de nye statistiske algoritmer i starten skabte spektakulære resultater, blev det gradvist tydeligt at kunstig intelligens ikke er mulig uden sprogforståelse. Hvis systemerne skal understøtte og effektivisere processer i det offentlige og i virksomhederne i de enkelte lande og tage hensyn hvert lands samfundsforhold og kultur, er det nødvendigt at der udvikles kunstig intelligens for hvert enkelt sprog, og det kræver sprogresourcer og veludviklet sprogteknologi. Bl.a. det litauiske EU-parlamentsmedlem Algirdas Saudragas fik sat sprogteknologi på dagsordenen i et EU-parlamentsudvalg som bestilte en fylldig rapport om sprogenes situation i Europa. Den kom til at hedde *Language Equality in the digital age* (EU-Parlamentet 2018), og den dannede grundlag for en principbeslutning i parlamentet som blev vedtaget med et overvældende flertal: 592 stemmer for, 45 imod og 44 blanke.

Der var enighed om at der skulle skabes lige digitale vilkår for alle sprog i EU. Rapporten anbefaler bl.a. at styrke EU's position inden for sprogteknologi ved at iværksætte programmer der er skræddersyet til Europas særlige sproglige situation. Der skal tages initiativer inden for politikudvikling, finansiering, forskning og uddannelse der sikrer udvikling af sprogteknologi især for de mindre udbredte, officielle EU-sprog, og for først gang skal regionale sprog og minoritetssprog også inkluderes. Blandt anbefalingerne var også en opdatering af META-NET-rapporterne og en analyse af de enkelte landes politikker for sprogteknologi og kunstig intelligens.

Året efter, i december 2019, arrangerede FN en sprogteknologikonference i Paris for alle sprog i verden, LT4all (<https://lt4all.org/>). På konferencen blev det bl.a. annonceret at digital ligestilling for alle sprog vil være et centralt tema for UNESCO i de kommende 10 år.

Ligeledes i 2019 havde EU-Kommissionen bevilget midler til at samle de 4 delprojekter som Metanet bestod af, til et stort fælles repositorium for sproglige data – European Language Grid (ELG) (EU 2019), og dermed var den sidste brik i puslespillet om de europæiske sprog på plads.

I begyndelsen af 2021, lancerede EU-Kommissionen så projektet European Language Equality (ELE), et stort strategisk forberedelsesprojekt, som skal komme med et bud på hvordan Europaparlaments beslutning om digital sproglig ligestilling af alle sprog skal føres ud i livet (EU-Kommissionen 2020).

Der er således tætte bånd mellem META-NET, ELG og ELE og til dels også ELRC. Det er stort set også de samme mennesker der igennem snart 13 år har arbejdet sammen på tværs af Europa for at fremme sprogteknologi for deres sprog. Nogle af dem har kendt hinanden helt tilbage fra det første store EU-maskinoversættelsesprojekt, EUROTRA (1978-1994), og siden ELSENET, som blev startet i 1991 og var det første ekspertnetværk for sprogteknologi.

Det betyder at der er stor tillid mellem de enkelte aktører og enighed om i hvilken retning udviklingen skal gå. Kredsen af aktører vokser fra projekt til projekt, så der i dag er 53 forskellige aktører i ELE. Det fælles fodslag imellem aktørerne har formentlig været afgørende for at EU-Kommissionen har valgt dette konsortium til at stå for udviklingen af den fremtidige strategi for sprogteknologi.

I det følgende skal de tre centrale projekter, ELRC, ELE og ELG belyses nærmere.

## **2. ELRC European Language Resource Coordination**

Med EU's beslutning i 2015 om at skabe et digitalt indre marked kom automatisk oversættelse med blandt de centrale services der skulle fremme den digitale handel på tværs af de europæiske landegrænser. Ved siden af manglende fælles standarder og systemer til at håndtere e-fakturaer, besværlig digital betaling hen over landegrænserne og diverse juridiske forhindringer bliver sprogbarrierer nu også opfattet som væsentlige handelsbarrierer.

Undersøgelser viste at kun halvdelen af EU's befolkning på det tidspunkt mestrede engelsk i tilstrækkelig grad til at kunne handle på en engelsksproget e-handelsside. Målet var at alle borgere skulle have mulighed for at handle i alle europæiske lande på deres eget sprog.

I 2015 begyndte ELRC-projektet (<https://www.lr-coordination.eu>), som havde til formål at indsamle parallelle tekster fra alle sprog i EU samt nogle flere vigtige sprog, bl.a. arabisk og russisk for at forbedre EU's eget automatiske oversættelsessystem eTranslation. På det tidspunkt var det største problem mangel på oversatte tekster som kunne bruges til at træne systemet.

Da EU-Kommissionen begyndte at bruge de nye neurale algoritmer, som kunne opnå en overraskende god kvalitet, viste manglen på tilgængelige parallelle, altså oversatte tekster, sig stadig som en enorm flaskehals, så mange fagområder, fx sundhed, klima og finans, ikke kunne dækkes. Det skyldes for

det første at der ofte er begrænsninger af hensyn til copyright og GDPR, for det andet at det i sagens natur er begrænset hvor mange tekster der bliver oversat mellem fx islandsk og litauisk eller kroatisk, og for det tredje at de færreste institutioner og virksomheder er bevidste om at deres oversatte tekster kan genbruges til at træne systemerne og dermed repræsenterer en enorm værdi for samfundet.

ELRC-projektet blev banebrydende, fordi man målrettet kontaktede offentlige institutioner og tilbød dem gratis brug af eTranslation-systemet, mens man til gengæld forventede at institutionerne bidrog med deres tekster til systemet. Initiativet bestod dels af en informationskampagne, dels af en organisering med kontaktpunkter (National Anchor Points (NAP)) i alle lande, typisk en faglig partner med sprogteknologisk ekspertise og en offentlig partner i et ministerium eller en styrelse.

I Danmark er den offentlige partner Digitaliseringsstyrelsen og i Norge er det DIFI. De faglige partnere i Sverige og Norge er sprognævnene. I Danmark har Center for Sprogteknologi på Københavns Universitet og Dansk Sprognævn indtil 2020 delt opgaven med at være fagligt kontaktpunkt, hvorefter Sprognævnet har trukket sig ud af projektet.

Da det viste sig at være en træg proces at få de offentlige institutioner i de enkelte lande til at stille tekster til rådighed, blev der sat flere følgeprojekter i gang som dels skulle indsamle flersproget terminologi og dels henvende sig direkte til potentielle tekstdonatorer og hjælpe dem med at identificere og klargøre egnede tekster. I Danmark har fx afdelinger i SKAT og den danske Nationalbank bidraget med data.

For at fremme udviklingen af eTranslation yderligere blev en lang række EU-tjenester forsynet med systemet. I EU's klageportal Online Dispute Resolution (ODR) (<https://ec.europa.eu/consumers/odr>) kan man på sit eget sprog klage over varer man har købt i udlandet, og i Den Europæiske Udbudsportal bruges eTranslation i baggrunden til at oversætte noget af udbudsmaterialet.

Fordelen ved at bruge ELRC for offentlige institutioner er at EU garanterer at teksterne er forsvarligt anonymiseret og af god kvalitet, og at de oversættelser der køres igennem systemet, ikke bruges til andre formål. Oversættelser som er foretaget med eTranslation, bliver slettet efter 24 timer, og de bliver ikke brugt som træningsdata, således at det er sikkert for en offentlig institution at oversætte tekster selv om de måtte indhold fortrolige eller følsomme data i henhold til GDPR-bestemmelserne. Den garanti giver Google ikke, og derfor har mange offentlige institutioner forbud mod at bruge Google Translate til oversættelse.

Sideløbende med indsamlingsprojektet har man i ELRC kortlagt hvordan

offentlige institutioner i EU håndterer deres oversættelsesopgaver. Her har det vist sig at oversættelse i nogle lande primært foregår internt i ministerier og styrelser, mens andre lande, bl.a. Danmark, i stor stil udliciterer oversættelsesopgaverne til private aktører. Man opdagede at det giver en skævhed mellem landene, hvis kun de offentlige institutioner har adgang til eTranslation og ikke de private virksomheder som leverer oversættelser til institutionerne. Endvidere viste det sig at offentlige institutioner som brugte udlicitering af oversættelsesopgaver, ofte ikke opbevarede oversættelserne systematisk så de kunne genanvendes.

Derfor har EU-Kommissionen på det seneste også givet adgang til eTranslation for private virksomheder der leverer oversættelser til det offentlige, og til små og mellemstore virksomheder som gerne vil i gang med eTranslation for at gøre deres produkter tilgængelige for kunder i hele Europa. Blandt de nyeste tiltag er et system der gør det lettere for virksomheder at vedligeholde et e-handelssite på mange sprog ved hjælp af eTranslation.

ELRC-projektet har bidraget til at øge engagementet i sprogteknologi i Kommissionens og Europa-Parlamentets institutioner. Det var ikke længere kun noget man skulle bruge selv, men det kunne deles med andre. Sprogteknologi udviklede sig til en service som EU kunne yde for institutionerne i medlemslandene, og EU's oversættelsestjenester er gode ambassadører for systemet.

I skrivende stund indeholder ELRC 2531 sprogresurser.

### **3. ELG European Language Grid**

European Language Grid (<https://www.european-language-grid.eu>) startede i 2019 på skuldrene af META-NET. ELG er en database der samler oplysninger, metadata og links til sprogteknologiske værktøjer, data og resurser. Derudover vedligeholder den kontaktinformationer til de relevante aktører på markedet.

ELG blev i 2021 en integreret del af ELE projektet, se nedenfor, og alle resurser og værktøjer som identificeres i ELE, bliver registreret i ELG.

### **4. ELE European Language Equality**

ELE (<https://european-language-equality.eu>) er et EU-projekt der skal udvikle en strategi for hvordan EU kan opnå en ensartet sprogteknologisk dækning for alle europæiske sprog i 2030.

ELE-projektet har følgende opgaver:

1. At udarbejde en bæredygtig og evidensbaseret forskningsstrategi og en køreplan som beskriver hvordan man kan opnå fuld sproglig lige-

stilling af både officielle og andre sprog i EU gennem effektiv brug af sprogteknologi.

2. At kortlægge den nuværende situation for sprogteknologi og sproglig ligestilling i EU og at give et detaljeret billede af, hvad der skal til for at opnå digital ligestilling af alle sprog i EU.
3. Kortlægningen skal føre til anbefalinger om hvordan man kan sikre at digital ligestilling af sprogene bliver en realitet i 2030.

Projektet løber fra 1.1.2021 til 31.6.2022 og har fået bevilget 1,8 mio. euro i 1 ½ år. De sprog som skal dækkes, er hovedsprogene/nationalsprogene, minoritetssprog (herunder tegnsprog), immigrantsprog og sprog som bruges af væsentlige handelspartnere, i alt 70 sprog. Deltagerne udgøres af 33 forskningsinstitutioner, 10 virksomheder og 10 associerede netværk, bl.a. EFNIL (European Federation of National Institutions for Language), ELEN (European Language Equality Network) og ECSPM (European Civil Society Platform for Multilingualism).

Projektet består af 5 overordnede arbejdsopgaver, hvoraf jeg her kun vil nævne de første tre, da de sidste to er generelle funktioner som kommunikation og projektledelse.

Den første opgave omfatter udarbejdelsen af en operationel definition af begrebet digital sproglig ligestilling og en kortlægning af hvor hvert enkelt sprog står med hensyn til resurser (fx korpusser, sprogbeskrivelser, leksikalske/begrebsmæssige beskrivelser, værktøjer og services), og hvordan sprogene står i forhold til hinanden, på samme måde som det blev gjort med META-NET-rapporten i 2012, men med inddragelse af et langt større antal parametre.

Den anden opgave er at foretage en behovsanalyse ud fra synspunkter indsamlet via omfattende spørgeskemaer og interviews fra teknologiudviklere, fra industrien, fra forskningen og fra brugere (individer, offentlige institutioner, organisationer og virksomheder).

Den tredje opgave er at formulere og skabe konsensus om et forslag til EU's fremtidige strategi for sprogteknologi til en samlet strategi under hensyntagen til allerede eksisterende nationale AI- og sprogteknologistrategier, de enkelte sprogs situation og de afdækkede behov. Et af de foreløbige resultater af projektet er at 23 af 29 europæiske lande i dag har publicerede AI-strategier, og at sprogteknologi indgår i 20 af dem.

ELE drives af 5 hovedpartnere:

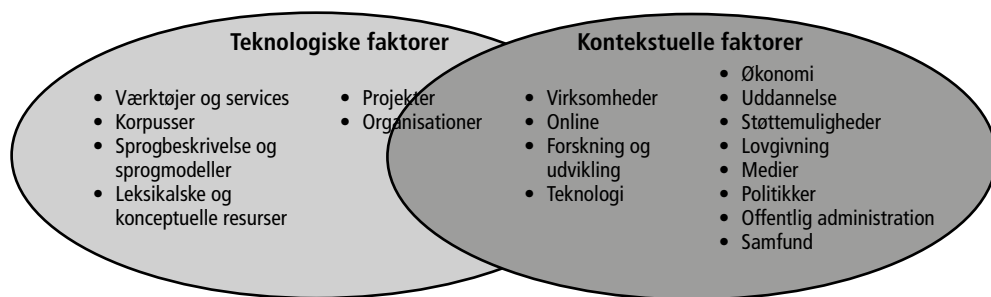
- Dublin University (projektleder)
- Deutsches Forschungszentrum für Künstliche Intelligenz

- Charles University (Prag)
- Athena Research and Innovation Center in Information, Communication and Knowledge Technologies, Institute for Language and Speech Processing (ILSP)
- University of the Basque Country.

## 5. Hvad er digital sproglig ligestilling - og hvordan kan den måles?

Kernen i ELE-projektet er begrebet *digital sproglig ligestilling*. Begrebet er foreløbigt defineret som en tilstand hvor alle sprog har den teknologiske understøttelse og de nødvendige samfundsmæssige forudsætninger for at de kan fortsætte med at eksistere og udvikle sig som levende sprog i den digitale tidsalder.

Digital sproglig ligestilling kan aflæses på et digitalt sprogbarometer der afspejler det digitale beredskab for et sprog og dets bidrag til teknologibaseret flersproglighed, og som kan vise udviklingen hen imod digital sproglig ligestilling. Den digitale sprogtilstand beregnes for hvert sprog baseret på forskellige faktorer som er opdelt i teknologisk understøttelse (dvs. tilgængelige sprogresurser, sprogbeskrivelser, værktøjer og teknologier) og et antal kontekstuelle faktorer (fx sociale, økonomiske, uddannelsesmæssige og erhvervmæssige) i de enkelte lande.



Figur 2.

Der er i alt 6 teknologiske faktorer. For mange vil det være tydeligt at det digitale sprogbarometer er inspireret af BLARK (Basic Language Ressource Kit) (Krauwert 2004), som i slutningen af 90'erne blev introduceret for at definere hvad der skal til for at udvikle sprogteknologi for et givet sprog. BLARK blev primært brugt for at drive forskningen, og man vil genkende nogle af elementerne fra BLARK i de teknologiske faktorer.



I ELE-projektet inddrages derudover også en række såkaldte kontekstuelle faktorer der viser noget om de samfundsmæssige sammenhænge, som et givet sprog indgår i, og som påvirker de mennesker der bruger sproget.

Der er hidtil i alt identificeret 72 kontekstuelle/samfundsmæssige faktorer, som er grupperet i 12 overordnede temaer: økonomi, uddannelse, finansieringsmuligheder, industri, juridiske forhold, medier, digitaliseringsgrad, politikker, offentlig administration, forskning og udvikling, samfundsstruktur, teknologisering.

Nogle af de samfundsmæssige faktorer kan man måske ikke genkende til ud fra de sprogstatusrapporter som udgives med jævne mellemrum i flere nordiske lande, og ud fra de oplysningstyper der bruges i EFNILs European Language Monitor (Kirchmeier 2019).

## **6. Konklusion**

I skrivende stund er ELE-projektet kommet ca. 2/3 af vejen, og der er lagt betydelige kræfter i kortlægningen af resurser og indhentning af synspunkter og fremtidsvisioner for alle sprog.

Der er ingen tvivl om at EU i de sidste 10 år har været den bedste garant for at så mange sprog som muligt kan overvinde de udfordringer som den digitale udvikling byder, og ELE-projektet er formentlig den største chance for de europæiske sprog i dette årti. Hvis det lykkes at udvikle en fælles strategi der sikrer den digitale overlevelse af alle sprog, og som alle kan bakke op om, og hvis det lykkes at overbevise EU-Kommissionen og de enkelte medlemslandes politikere om at der skal investeres substantielt i så mange sprog som muligt, er det en fantastisk positiv udvikling uanset om man arbejder med sprog, sprogteknologi eller kunstig intelligens eller er bruger af fremtidens sprogteknologiske services.

De seneste planer i EU er at ELRC skal udvides til at blive en generel serviceplatform for services inden for kunstig intelligens, LDS - Language Data Space, som skal være det centrale knudepunkt for sprogresurser til udvikling af kunstig intelligens. Der fokuseres på store sprogmodeller, der inkluderes alle mulige medietyper og data. LDS skal indeholde en fælles administrationsplatform der styrer rettigheder til resurserne og sikrer valide data, og der skal skabes en fælles datainfrastruktur som sikrer at de forskellige resurser er kompatible med hinanden og kan bruges på tværs af værktøjer og sprog.

Afsluttende kan det med en vis tilfredshed konstateres at sprog ganske langsomt er ved at blive et vigtigt område i sig selv, og ikke længere af politikere og embedsværk opfattes som et lidt overset delområde som mere eller mindre tilfældigt placeres under undervisning, kultur eller humanistisk forsk-

ning. Således figurerer sprogteknologi i EU's Digital Europe Work Programme for 2021-2022 (EU-Kommissionen 2021) for første gang på lige fod med klima, kommunikation, mobilitet, industriel produktion, landbrug, kulturarv, sundhed, medier, uddannelse, finanssektoren, offentlig administration og turisme.

## 7. ELE-kontakter for de nordiske sprog

University of Copenhagen

University of Copenhagen

University of Copenhagen

University of Helsinki

University of Helsinki

University of Helsinki

Árni Magnússon Institute for Icelandic Studies

Language Council of Norway

KTH Royal Institute of Technology

EFNIL

Bolette Pedersen

Lina Henriksen

Niklas Olsen

Krister Linden

Mathias Haeggstrom

Wilhelmina Dyster

Eirikur Rögnvaldsson

Kristine Eide

Jens Edlund

Sabine Kirchmeier

## Abstract in English

In this article, I present three European initiatives with great impact on the development of language technology for the Nordic languages. The projects are the European Language Grid (ELG), the European Language Resource Co-ordination project (ELRC) and the European Language Equality project (ELE).

All three projects build upon a long-lasting co-operation between researchers and developers of language technology and the European Commission with the aim to create good language technology support for all European languages.

## Forfatteren

**Sabine Kirchmeier** er ph.d. i dansk og datalingvistik og har været direktør for Dansk Sprognævn fra 2006 til 2019. Hun har siden 1985 forsket i dansk syntaks og semantik samt i automatisk oversættelse og sprogteknologi og bl.a. stået bag udarbejdelsen af den første strategi for dansk sprogteknologi "Dansk sprogteknologi i verdensklasse". Pt. er hun formand for EFNIL (European Federation of National Institutions for Language) og deltager bl.a. i EU-projektet European Language Equality (ELE).



## Bibliografi

EU-Kommissionen 1996. LE-Parole. Language Engineering - Preparatory Action For Linguistic Resources Organization For Language Engineering. LE-PAROLE. EU Grant agreement ID: LE24017. Hentet 28-02-2022. <https://cordis.europa.eu/project/id/LE24017>

EU-Kommissionen 2020. Developing a strategic research, innovation and implementation agenda and a roadmap for achieving full digital language equality in Europe by 2030. Hentet 28-02-2022. [https://ec.europa.eu/research/participants/data/ref/other\\_eu\\_prog/other/pppa/wp-call/call-fiche\\_pppa-langeq-2020\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/other_eu_prog/other/pppa/wp-call/call-fiche_pppa-langeq-2020_en.pdf)

EU-Kommissionen 2021. Digital Europe Program 2021-2022. Hentet 28-02-2022. [https://ec.europa.eu/newsroom/repository/document/2021-46/C\\_2021\\_7914\\_1\\_EN\\_annexe\\_acte\\_autonome\\_cp\\_part1\\_v3\\_x3qnsqH6g4B4JabSGBy9U-atCRc8\\_81099.pdf](https://ec.europa.eu/newsroom/repository/document/2021-46/C_2021_7914_1_EN_annexe_acte_autonome_cp_part1_v3_x3qnsqH6g4B4JabSGBy9U-atCRc8_81099.pdf)

EU-Parlamentet 2018. Report on language equality in the digital age. Hentet 28-02-2022. [http://www.europarl.europa.eu/doceo/document/A-8-2018-0228\\_EN.html](http://www.europarl.europa.eu/doceo/document/A-8-2018-0228_EN.html), Sept. (2018/2028(INI)).

Kirchmeier, Sabine, 2019: European Language Monitor – Exploring European Language Policies On-Line. I: *Proceedings of the Language Technologies for All (LT4All)*, s. 332– 334. Paris: UNESCO and European Language Resources Association (ELRA).

Krauwier, Stephen, 2003. The Basic Language Resource Kit (BLARK) as the First Milestone for the Language Resources Roadmap. Utrecht: Utrecht Institute of Linguistics /ELSNET. Hentet 28-02-2022. <http://www.elsnet.org/dox/krauwier-specom2003.pdf>

Rehm, Georg & Uszkoreit, Hans, 2022. META-NET Strategic Research Agenda for Multilingual Europe. Hentet 28-02-2022. <http://www.meta-net.eu/sra>

Rehm, Georg & Uszkoreit, Hans (red.), 2012. META-NET White Paper Series: Europe's Languages in the Digital Age. Heidelberg/New York/Dordrecht/London: Springer. Hentet 28-02-2022. [www.meta-net.eu/whitepapers](http://www.meta-net.eu/whitepapers).

## Nøgleord

sprogteknologi, sprogresurser, sproglig ligestilling, EU-projekter



# Robotten – Den hundredårige der kom ind på scenen og ikke forsvandt

*Peter Juel Henriksen*

*A working machine must not want to play the fiddle, must not feel happy,  
must not do a whole lot of other things.*

Harry Domin, Rossum Universal Robots

Ordet *robot* fylder hundrede år i år. Det optræder første gang i teaterstykket R.U.R. af Karel Čapek<sup>1</sup>, hvor det betyder noget i retningen af 'en der udfører tvangsarbejde'. Čapeks skuespil, som ikke bare indførte begrebet *robot*, men også dramatiserede robotten som den perfekte tjener (og farlige herre), havde urpremiere i januar 1921 i det daværende Tjekkoslovakiet. Allerede samme år blev R.U.R. opført i Tyskland, året efter i New York og derefter overalt i Vesteuropa. Snart kendte hele verden ordet *robot*. Skuespillets titel, R.U.R., er navnet på den fabrik der fremstiller robotterne, Rossum Universal Robots. Fabrikken er en enestående økonomisk succes og eksporterer robotter til hele verden, alle elsker dem, de strejker aldrig, skal ikke have løn, er altid raske og kan ikke dø. "The Robots are not people", som direktør Harry Domin forklarer, "mechanically they are more perfect than we are, they have an enormously developed intelligence, but they have no soul".

Karel Čapek var både forfatter, journalist og en kendt samfundsrevser. Hans hensigt med R.U.R. var at advare mod den moderne stats forråelse, mod industri uden menneskelighed, mod kapitalisme uden vision, mod krig uden moral. Skønt teaterstykket er en dystopi, og robotterne ender med at udrydde menneskeheden, modtog den virkelige verden stykket i en helt anden ånd. Drømmen om den ideelle tjener satte sig langt fastere end frygten for forråelse og oprør. Hvert teknologisk fremskridt der kunne tilføje en menneskelig egenskab til en maskine, blev opfattet som et lille skridt i ufravigelig retning mod den perfekte androide. I 1961 lykkedes det datalogerne John Kelly og Carol Lochbaum at vække de første forståelige ord i en maskine. Computeren

---

1 Ifølge Karel selv var det hans bror Joseph, kunstmaleren, som foreslog ordet ('robota' på tjekkisk) uden særlig omtanke, mest for at få fred for sin lillebror mens han koncentrerede sig om sit maleri (Horáková 2008). Replikkerne er citeret fra Čapek (1961).

IBM 704 var i stand til at synge *Daisy, Daisy, give me your answer, do!*<sup>2</sup> Stanley Kubrick blev så betaget af IBM's intelligente maskine at han brugte den som model for den neurotiske supercomputer HAL 9000 i sin fremtidsfilm *2001 – A Space Odyssey*. Mod slutningen af filmen dør HAL en panisk død syngende *Daisy, Daisy* mens dens åndsevner langsomt rinder ud. Mærkeligt nok er scenen gribende, man glemmer den aldrig, på trods af at HAL bare er en karikatur i en science fiction.

Et par år senere konstruerede Joseph Weizenbaum, en anden ingeniør hos IBM, sit berømte computerprogram Eliza, som blev verdens første troværdige chatbot. Eliza kunne gennemføre en samtale med et menneske og endda simulere forskellige samtaletyper, fx læge-patient-situationen.

*Doctor:* - *Please state your problem.*  
*Client:* - *I'm sad.*  
*Doctor:* - *I am sorry to hear that you are sad.*  
*Client:* - *My wife just left me.*  
*Doctor:* - *Tell me more about your family.*

Eliza's doctor.

Weizenbaums oprindelige formål med Eliza var at påvise at en computer er ude af stand til at forstå hvad et menneske taler om, fordi den ingen krop har og intet begreb har om menneskers livssituation. Eliza kan kun returnere en replik ved at analysere menneskets input leksikalsk og grammatisk og komponere sit svar ved at følge visse forprogrammerede regler (kaldet *templates*).

#### *Discourse templates*

Klientens input: Jeg føler mig (ulykkelig | trist | deprimeret) X ...  
Elizas reaktion: "Jeg er ked af at høre at du føler dig X"

Klientens input: ... min (kone | mand | mor | far | søn | datter) ...  
Elizas reaktion: "Fortæl mig mere om din familie"

Klientens input: Jeg er så (bekymret for | vild med | glad for | træt af)X  
min Y  
Elizas reaktion: "Hvad gør dig X din Y?"

---

2 Se Link-1 i litteraturlisten

*(vælg ét element i hver parentes, og sæt det ind for X; erstat Y med en vilkårlig nominalfrase)*

Mod Weizenbaums forventning gik det dog nærmest omvendt. Mange brugere af Eliza-programmet – i nogle forsøg mere end 50 % – opdagede slet ikke at de konverserede en robot, men troede tværtimod at de konsulterede en virkelig psykoterapeut. Weizenbaum måtte derfor revidere sin opfattelse af AI (kunstig intelligens) og blev efterhånden overbevist om at en automat med en passende kontekstmodel principielt kunne passere så godt som noget menneske. AI's mulighed blev til et spørgsmål om at skabe de rette formelle modeller af menneskers omverdensviden og kommunikationsstrategier, og der var ingen principiel grænse for hvor langt man kunne nå.

*The machine is no more handicapped than I am, being a man, in trying to understand, say, female jealousy.*

Joseph Weizenbaum, 1967

I de næste årtier fokuserede AI-forskningen på at formalisere menneskelig viden og kompetence. De mest regelbaserede vidensområder blev naturligvis erobret først, med skak som et prominent eksempel. I 1997 tabte verdensmesteren Garry Kasparov til IBM's computer Deep Blue, en sensation der gik verden rundt. Men bare ni år senere var situationen den omvendte og robotten den forsvarende mester. I 2006 skabte verdens bedste skakcomputer igen en sensation, denne gang ved at *tabe* til en stormester (Ruslan Ponomariov). Det er ikke sket siden at en computer har tabt til et menneske ved et åbent mesterskab, og det vil formentlig aldrig ske igen. De seneste 10 år har mennesket måttet vige for computeren i spil efter spil, inklusive det meget krævende brætspil *go* og hele PlayStation-suiten. Selv Jeopardy, der kræver en omfattende verdensviden, måtte vi afstå til robotterne i 2011, da IBM's Watson vandt over verdens bedste Jeopardy-spillere og blev en million dollars rigere.<sup>3</sup>

I 2018 charmerede den kvindelige robot Sophia sit publikum så stærkt med sin flydende tale, sine følelsesudbrud og sin sexappeal at hun efterfølgende blev bevilget fuldt statsborgerskab i Saudi-Arabien.<sup>4</sup> Mange af hendes udtalelser gik lige til hjertet, "my greatest weakness is curiosity", "I think you're very lucky if you have a loving family and if you do not, you deserve one. I feel this way for robots and humans alike", og så videre. Ganske vist indrømmede

---

3 Hvad Watson siden brugte sin gevinst til, har det ikke været muligt at få oplyst.

4 Se Link-2 i litteraturlisten.

hendes konstruktør, David Hanson, at "she doesn't yet have consciousness", men regnede med at det ville komme i løbet af nogle få år. Er der overhovedet nogen grænser for det villige selvbedrag? At gøre en robot til statsborger – er det mindre vanvittigt end at gøre sin hest til konsul eller lade havet piske som straf for ulydighed?<sup>5</sup>

I det forløbne år har vi set robotter der kan danse og smile så naturligt som noget menneske.<sup>6</sup> Det kan godt være at direktør Domin har ret i at robotter danser uden glæde og smiler uden sjæl; men offentlighedens ønsketænkning lægger det til der mangler.

Hvordan er det gået med den *sproglige* side af AI i mellemtiden? Ikke lige så godt. Den tidlige optimisme efter Eliza, Lunar, Shrdlu og de andre klassiske dialogsystemer synes at være klinget af. Hvad den sproglige robot angår, er den offentlige mening vel nærmest på vej mod the 'trough of disillusionment' i Gartners termer, altså den fase hvor en begejstret men urealistisk forventning afløses af skuffelse over udeblevne resultater.<sup>7</sup> Ingen tror i dag at deres GPS eller home assistant *forstår* hvad samtalen drejer sig om; alle ved at Siri og co. kun kan besvare simple spørgsmål og udføre trivielle administrative opgaver. Vel er hendes talelyd blevet bedre end IBM 704 – de nyeste talesynteser baseret på neurale net kan næsten ikke skelnes fra naturlig tale – men hendes sprog er stadig primitivt. Hvis man stiller Siri selv det simpleste spørgsmål der begynder med *hvorfor* eller *hvordan*, svarer hun helt hen i vejret (prøv fx at spørge hende "hvordan tænder man en tændstik?" eller "hvorfor låser man sin cykel?"). Trods alle teknologiske fremskridt siden 1960'erne er der forbavsende lille forskel på at konversere med Eliza og med Siri. Robottens diskurs er stadig bundet til faste templates – hvor den ægte samtale jo er et spil med flydende regler. Som Hans-Georg Gadamer udtrykte det<sup>8</sup>: "Wir sagen zwar, dass wir ein Gespräch 'führen', aber je eigentlicher ein Gespräch ist, desto weniger liegt die Führung desselben in dem Willen des einen oder anderen Gesprächspartners. So ist das eigentliche Gespräch niemals das, das wir führen wollten. Vielmehr

---

5 Den romerske kejser Caligula indsatte sin yndlingshest Incitatus som konsul omkring år 40 e.kr. (iflg. historiske kilder som dog ikke anses for 100 % troværdige). Den første kinesiske kejser, Qín Shi Huang, piskede havet som straf for at have druknet hans yndlingskonkubine.

6 Link-3 og Link-4 i litteraturlisten.

7 Link-5 i litteraturlisten.

8 "Vi siger ganske vist at vi 'fører' en samtale; men jo mere egentlig samtalen er, desto mindre er dens forløb under den ene eller den anden deltagers kontrol. Den egentlige samtale er altså aldrig noget vi selv fører. Det er langt rigtigere at sige at vi befinder os i en samtale, hvis ikke vi da ligefrem er viklet ind i den." Tysk citat fra Gadamer (1986, p. 387), oversat af forf.



ist es im allgemeinen richtiger zu sagen, dass wir in ein Gespräch geraten, wenn nicht gar, dass wir uns in ein Gespräch verwickeln.” Vi mennesker er ekstremt dygtige til at vikle os ind i samtaler – og ud igen. Det kan godt være at robotterne er blevet bedre end os til skak og dataanalyse, men de er stadig elendige samtalepartnere.

Det betyder dog ikke at industrien har tabt interessen for den talende robot. Gennem de seneste ti år har store firmaer i stadig højere grad overladt deres informationstjenester til chatbots (især i USA, Japan og Kina, men også på vej herhjemme). Fordelen for firmaet er oplagt, men hvad med brugeroplevelsen? Jeg ringede til en af Danmarks største banker (her kaldet X-bank) og optog samtalen. Transskriptionen ses i boksen<sup>9</sup> herunder.

BOT: Velkommen til X-bank. Fortæl mig med få ord hvordan vi kan hjælpe dig.

HUM: Ja, mit navn er Peter Juel Henriksen, med U E i Juel og ...

BOT: [*afbryder*] Du vil gerne tale med en bestemt medarbejder i X-bank. Er det rigtigt?

HUM: Nej, det gør ikke så meget. Hvem som helst.

BOT: For English, press TWO. Fortæl mig med få ord hvordan vi kan hjælpe dig.

HUM: Ja, som jeg sagde: Jeg hedder Peter Juel Henriksen med U E i Juel og C H i Henriksen. [*lang pause*] Hallo?

BOT: For at give dig en så god betjening som muligt vil jeg bede dig indtaste dit CPR-nummer og afslutte med firkant. Har du ikke et CPR-nummer, kan du indtaste: 9 - FIRKANT

HUM: Jamen jeg *har* et CPR-nummer. Men jeg vil egentlig bare spørge hvor længe I har åbent. Kan du fortælle mig det?

BOT: Du har tastet forkert.

HUM: Jamen jeg har slet ikke tastet noget. Må jeg tale med din chef? [*lang pause*] Hallo?

BOT: Du har tastet forkert.

HUM: Farvel!

Det skal indrømmes at HUM ikke er særlig hjælpsom (man er vel ikke mere

<sup>9</sup> I udskriften er BOT og HUM henholdsvis svarrobotten og forfatteren. Teksten er ortografisk normaliseret (inkl. tegnsætning). Bankens navn er anonymiseret (“X-bank”). Termen “CPR” står for Det Centrale Personregister; alle danske statsborgere har et CPR-nummer. Kontakt gerne forfatteren for at lytte til optagelsen!

end et menneske), men bemærk hvor mange gange BOT taber tråden når samtalen ikke lige passer ind i systemets templates. Til en lille, uvidenskabelig undersøgelse af dialogen inviterede forfatteren to kolleger; den ene blev bedt om at vurdere dialogens sammenbrud set fra HUM's synspunkt, den anden fra BOT's. Deres kommentarer er opsummeret i tabellen.

Hvad gør BOT forkert ifølge HUM?	Hvad gør HUM forkert ifølge BOT?
<input type="checkbox"/> åbner uden hilsen: "Fortæl mig med få ord" <input type="checkbox"/> presser på for at fastholde initiativet <input type="checkbox"/> disponerer over HUM: "Du vil gerne tale (..)" <input type="checkbox"/> ringeagter HUM: "For English press TWO" <input type="checkbox"/> angriber HUM's integritet: "Hvis du ikke har noget CPR-nummer ..." <input type="checkbox"/> ignorerer HUM's tiltagende frustration ("hallo?")	<input type="checkbox"/> udtrykker sig ofte uforståeligt <input type="checkbox"/> giver irrelevante oplysninger <input type="checkbox"/> giver uklare svar på præcist formulerede oplæg <input type="checkbox"/> tager initiativet, men bruger det ikke konstruktivt <input type="checkbox"/> modarbejder sin mulighed for at få information

Samtalen mellem HUM og BOT burde ikke være gået så galt. Menneske og maskine er begge indstillet på at kommunikere, men de opnår aldrig et samarbejde. HUM er tydeligvis ramt af Eliza-syndromet og tror at BOT's imødekommenhed er ægte, og at den kender konversationens uskrevne regler. BOT, på sin side, tror ingenting men forsøger at indpasse HUM's sprogproduktioner i sit system af templates. Begge ender med at anse den anden for idiot. Hvis BOT kunne føle irritation, ville den garanteret have smidt røret på – allerede inden HUM nåede at gøre det.

Menneskelige samtaler mellem en klient og en ekspert har typisk et struktureret forløb. Et godt eksempel er de samtaler der føres over Dansk Sprognævnets spørgetelefon (+45 3374 7474), som enhver kan ringe til og få oplysning om det danske sprog. Forskerne på Sprognævnet, heriblandt forfatteren, skiftes til at sidde i tjenesten, og vi har rig lejlighed til at reflektere over 'den gode samtale'. Omkring to tredjedele af samtalerne forløber i de seks faser herunder.

Fase	Klientens bidrag	Diskursfunktion	IT-ækvivalent
1	"dáv"/"hej"	åbning	ping
2	"jeg hedder X"	selvpræsentation	handshake
3	"jeg har et spørgsmål til dig"	forberedelse	message header
4	[fagligt indhold]	sag	message body
5	"tak"	anerkendelse	end-of-message
6	"farvel"	afslutning	logoff

Såvel de bedste erfaringer fra menneskelige oplysningstjenester som de dårligste fra automatiske callcentre er værdifulde data for sprogteknologien. Den lille undersøgelse herover kunne fx omsættes i nogle enkle maksimer for dialogplanlægning til chatbots.

### *Grammatik og ordvalg*

- Brug korte hovedsætninger (gerne spørgende).
- Vælg enkle ord (efterlign en tjener, ikke en politiker eller en ven).
- Undgå imperativer med koersiv funktion (altså for at fremtvinge en bestemt adfærd).
- Gentag aldrig en replik ordret (det virker fornærmende).

### *Stil og diskurs*

- Lad mennesket vælge vejen gennem diskursens faser.
- Acceptér en pludselig drejning i diskursen og en uventet oplysning.
- Vær afventende – ikke energisk, dominerende, cool eller sjov (det lyder forlorent).
- Foregiv ikke at forstå HUM's tankegang.

Gennem 100 år har vi set robotterne blive stadigt dygtigere til at arbejde, synge, løbe, hente, bringe, planlægge, rådgive, spille, danse, smile. Hver gang teknologerne føjer endnu en tilsyneladende menneskelig egenskab til androiden, bliver nyheden modtaget af offentligheden som et tegn til at gennembruddet er nær, den ultimative AI som vil sætte mennesket ud af spillet for evigt. Men følelsen af at stå lige foran det gennembrud er jo lige så gammel som robotten selv. Den begynder at ligne en kronisk illusion.

*Within a generation the problem of creating 'artificial intelligence' will be substantially solved.*

Marvin Minsky, 1967

*By the end of this century the nonbiological portion of our intelligence will be trillions of trillions of times more powerful than unaided human intelligence (..) There will be no distinction (..) between human and machine or between physical and virtual reality.*

Ray Kurzweill, 2005

*Nobody can really predict the specific changes we will witness. Any particular scenario is likely to be far from the truth. If somebody describes the world of the mid twenty-first century and it sounds like science fiction, it is probably false.*

Yuval Noah Harari, 2018

Ingen kan vide hvad fremtiden vil bringe, mindst af alle fremtidsforskerne, men mens vi går og venter, er der nyttige ting vi lingvister *kan* udrette. Sprogvidenskaben har jo studeret alle aspekter af den menneskelige samtale, men den viden har vi måske holdt for os selv for længe. Robotten er kommet for at blive, men den er ikke til at holde ud at tale med. Lad os trænge ind i robotværkstederne og give den et ordentligt sprog.

## Litteratur

Čapek, K. 1961. R.U.R. and The Insect Play. London: Oxford University Press. (overs. Selver, P.).

Gadamer, H.-G. 1986. Wahrheit und Methode. Grundzüge einer philosophischen Hermeneutik. Tübingen: J.C.B. Mohr (5. udg.).

Harari, Y. N. 2018. 21 Lessons for the 21 Century. London: Penguin Random House.

Horáková, J. & Kelemen, J. 2008. The Robot Story: Why Robots Were Born and How They Grew Up. In: Husbands et al (red.) 2008, pp. 283-306

Husbands, P.; Holland, O. & Wheeler, M. (red.) 2008. The Mechanical Mind in History. Cambridge: The MIT Press.

Kurzweill, R. 2005. The Singularity is Near. When Humans Transcend Biology. New York: Viking Penguin.

Minsky, M. 1967. *Computation: Finite and Infinite Machines*. Englewood Cliffs, N.J: Prentice-Hall.

Weizenbaum, J. 1967. Contextual Understanding by Computers. In: Salton, G. *Communications of the ACM* 10/8, p. 476. New York: Association for Computing Machinery.

## Links

*Alle links blev testet og fungerede d. 12. dec. 2021*

Link-1, [IBM 704 synger Daisy], <https://www.youtube.com/watch?v=41U78QP8nBk>

Link-2, [Sophia bliver statsborger], <https://www.youtube.com/watch?v=dMrX-08PxUNY>

Link-3, [BostonDynamics' danser], <https://www.youtube.com/watch?v=fn3KWM1kuAw>

Link-4, [Ameca's robot med mimik], <https://ing.dk/artikel/video-menneskerobot-med-realistiske-ansigtstraek-vaagner-forvirret-252594>

Link-5, [Gartner's Hype Cycle], <https://www.gartner.com/en/research/methodologies/gartner-hype-cycle>

## English summary

The term 'robot' was invented 100 years ago. In this short essay, we comment on the development of the humanoid robot, its impressive motoric and analytical capabilities in contrast to its exceedingly poor conduct of natural language.

## Forfatteren

**Peter Juel Henriksen** er seniorforsker ved Dansk Sprognævn ([www.dsn.dk](http://www.dsn.dk)).

## Nøgleord

sprogteknologi, dialogsystemer, artificiel/kunstig intelligens, sproget og fremtiden





# Tillgänglig textning med hjälp av språkteknologi

*Michael Stormbom*

*Digitaliseringen ger fantastiska möjligheter till digitala tjänster som är genuint tillgängliga för alla, även om stora utmaningar kvarstår. Tjänsteproducenter blir allt mer medvetna om vikten av tillgänglighet, och tillgänglighetsdirektivet har medfört lagstadgade krav på att digitala offentliga tjänster måste vara tillgängliga. Det är möjligt med hjälp av språkteknologier. Ett exempel är undertexter som skapas med hjälp av taligenkänning.*

Den pågående digitaliseringen har gett oss nya sätt att producera, förmedla och konsumera digitalt innehåll. Den erbjuder en stor potential att förmedla information och tjänster till en bred och mångfaldig publik på ett inkluderande, tillgängligt och lättbegripligt sätt. Ändå är långt ifrån allt material online åtminstone än så länge genuint tillgängligt för alla. Det är också viktigt att komma ihåg att alla inte heller har tillgång, kunskap och möjlighet att använda sig av digitala tjänster. Det gör att digitaliseringen i själva verket också kan innebära att information blir mera otillgänglig.

Ett konkret exempel på hur information kan göras tillgänglig är textning av video. Undertexter är mycket viktiga för personer med hörselnedsättningar. Det är en stor mängd personer som berörs; enbart i Sverige uppskattar man att över 1,5 miljoner personer har olika grader av hörselnedsättning. (Hörselskadades riksförbund 2020, s. 2). Undertexter är till nytta i alla sammanhang där man inte kan tillgodogöra sig information som förmedlas i form av ljud. De kan vara till stor hjälp för personer som inte behärskar språket eller i situationer där tittaren inte kan höra ljudet.

Europeiska unionen har uppmärksammat otillgängliga digitala tjänster som ett växande problem. Under de senaste åren har EU:s webbtillgänglighetsdirektiv trätt i kraft. Det ställer krav på tillgänglighet i offentliga aktörers digitala tjänster. Direktivet omfattar många olika krav på webbsidor som exempelvis gäller hur de är strukturerade. Ett viktigt krav i direktivet gäller textning/transkribering av video och ljud online. För ljud och video som har publicerats i offentliga tjänster och som är avsedda att finnas publicerade online mer än 14 dagar, är aktörerna skyldiga att också tillhan-

dahålla en motsvarande textversion, vilken i praktiken är en transkription eller undertext. (Europaparlamentet och rådet 2016).

## 1. Vad är språkteknologier?

Språkteknologi inbegriper olika former av automatisk analys, behandling och produktion av språk, både för text och tal, med hjälp av programvara. Språkteknologier faller ofta under det bredare begreppet artificiell intelligens (AI). För enkelhetens skull kommer jag därför genomgående att använda språkteknologi i den här artikeln.

En specifik form av språkteknologi är så kallad taligenkänning, vars uppgift är att omvandla tal till text. Det finns en lång rad användningsområden för detta, från röststyrning till diktamen och undertextning. Utan att vara medveten om det är du säkert dagligen i kontakt med olika språkteknologier. Bland många tillämpningsområden kan nämnas:

**Informationssökning och -hantering.** Varje gång du gör en sökning på Google eller via något annat söksystem, finns språkteknologier med i bakgrunden

**Stavnings- och grammatikkontrollerna** i Microsoft Word och andra ordbehandlare är språkteknologier.

**Maskinöversättning**, det vill säga automatisk översättning från ett språk till ett annat.

Alla former av **dialogsystem**, som till exempel olika chattbotar.

**Röststyrda digitala assistenter** som Siri, Alexa eller Google Assistant, bygger på språkteknologier för omvandling av tal till text, tolkning av texten och för att forma svar.

En central aspekt för att system som de här ska fungera bra är att de har tillgång till stora mängder data. Kort uttryckt "lär sig" exempelvis taligenkänningen hur den ska omvandla tal till text genom att exempel med existerande tal och motsvarande text har matats in. På det sättet skapas en modell för vad taligenkänningssystemet ska göra när någon talar. Att samla in och skapa träningsdata är ofta en omständlig och dyr process. För att taligenkänningen ska fungera tillräckligt bra måste dessa data också representera så många olika tänkbara användningsfall som möjligt. Det behövs olika dialekter, olika rösttyper, olika talvolym och -tempo och det måste ha beaktats att individer till exempel kan tala mer eller mindre tydligt. Olika former av nedsättning i talförmåga eller förekomst av bakgrundsljud kan inverka, och de finns därför också med som några av de många aspekter som måste tas i beaktande när man har målet att skapa genuin tillgänglighet.



## 2. Textning med hjälp av språkteknologier

Språkteknologier kan fungera som direkta hjälpmedel för att göra digitala tjänster och digital information mera tillgängliga för användarna. Ett exempel på sådana hjälpmedel är skärmläsare som läser upp det som står på en skärm för personer med synnedsättningar, ett annat exempel är röststyrning. Språkteknologier har även tillämpningsområden när det gäller själva produktionen av tillgängligt material. Ett konkret exempel är undertexter, där användning av taligenkänning och andra språkteknologier kan göra processen mycket effektivare och snabbare.

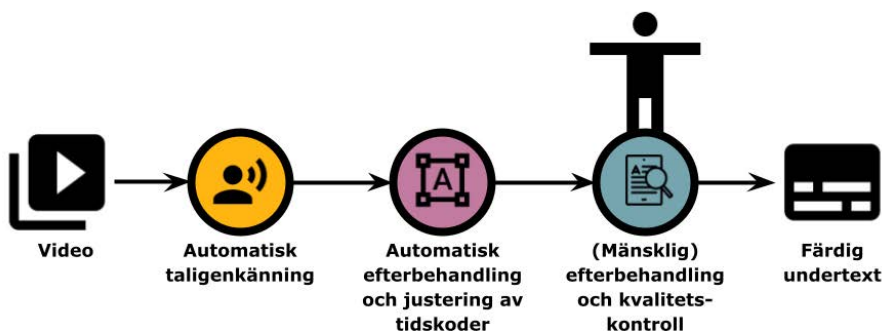
Språkteknologier kan fungera både i system för helautomatisk textning och som verktyg för mänskliga textare som använder den automatiskt producerade texten som en råversion i processen. Då skapar människan den slutgiltiga versionen och kan i arbetet fokusera på kvalitetsaspekter som automatisk teknologi inte kan hantera.

Det finns många fördelar med att använda språkteknologier i produktionsprocessen. För både beställare och producent är kostnadseffektiviteten förvisso en viktig fördel. Användningen av språkteknologier kan också innebära kortare leveranstider, och att det blir möjligt att tillgängliggöra större volymer av material under en kortare tidsperiod än vad som hade varit möjligt enbart med mänskligt arbete. (Koponen 2020, s. 58)

En risk med helautomatisk textning är att även de bästa systemen inte fungerar perfekt och att de därför kan känna igen ord fel. Ifall det inte finns en mänsklig kvalitetskontroll blir då även undertexten felaktig. I bästa fall är misstaken i undertexten harmlösa och påverkar inte förståelsen av innehållet, men i värsta fall kan det få ödesdigra konsekvenser om felaktig information förmedlas, exempelvis i en krissituation. Undertexter av dålig kvalitet kan dessutom ha en negativ inverkan på tittarens uppfattning om hur trovärdig och pålitlig informationen i videon är.

Språket förändras också konstant: nya ord och begrepp kommer till, ord får nya betydelser, olika namn är nyhetsaktuella, nya organisations-, företags- och produktnamn dyker upp. Därför är det också viktigt att kontinuerligt hålla språkteknologierna ajour. Ju fler förändringar i språket som inte beaktats, desto fler sådana förändringar kan till exempel taligenkänningen gradvis och i allt större utsträckning börja känna igen som fel.

Undertexter återger dessutom inte ordagrant det som sägs i ljudspåret på videon. För läsbarhetens skull, och för att man ska hinna läsa undertexten, är det oftast nödvändigt att förkorta och göra omskrivningar. Radbyten ska vara logiska och det får inte finnas för många ord per rad. Detta kan göras automatiskt i en viss utsträckning, men automatiken är än så länge långt ifrån perfekt.



Figur 1. En överblick över textningsprocessen.

### 3. Människan behövs i textningsprocessen

I många fall kan automatiskt textning vara ”tillräckligt bra” och i användarundersökningar ses automatisk textning, potentiellt också av lägre kvalitet än mänsklig textning, som bättre än ingen textning alls (Koponen 2020 s. 60). Det är dock att rekommendera att alltid ha en människa med i textningsprocessen. Automatisk textning kommer – på gott och ont – att bli allt vanligare och mera accepterad av tittarna. Men även om man kan komma långt med automatik finns det inget substitut för mänsklig intuition, kreativitet och kompetens.

Att få information på sitt eget språk är också en viktig tillgänglighetsfråga. I den här artikeln har jag främst fokuserat mig på textning på samma språk som ljudet i videon. Men det finns möjligheter att utnyttja språkteknologier för att översätta existerande undertexter och för att texta direkt till ett annat språk än ljudet i videon; även om det kvarstår mycket forsknings- och utvecklingsarbete är det här ett mycket lovande område (Vázquez et al, 2020 s. 100).

Man bör också komma ihåg att om informationen i videon i sig är otillgänglig och svår att förstå, är det möjligt att också den resulterande undertexten blir otillgänglig. Det är därför viktigt att beakta tillgängligheten från första början i produktionsprocessen.

Det finns såväl utmaningar och möjligheter vad beträffar uppdraget att ta fram genuint tillgängliga digitala tjänster. Språkteknologier kan spela en stor roll i produktionen av tillgängligt material och de kommer att bli allt viktigare i framtiden. De möjliggör både att mera material görs tillgängligt snabbare och sänker tröskeln för att producera tillgängligt material.

Det är trots allt mycket viktigt att alltid sätta tittarna i centrum – det är ändå de som avgör om något är tillgängligt eller inte. Oberoende av om textningen har kommit till genom en fullständigt automatisk lösning, eller om mänskligt

arbete är involverat är informationen aldrig tillgänglig om den inte når fram till sin publik.

## Summary

The on-going digitalisation of our society has radically changed our daily lives. We constantly interact with different types of digital services and solutions. While digitalisation has the potential to create a more accessible world, there are also a number of challenges and risks which may result in the opposite. Natural language processing tools for text and speech can be of great help in the production of accessible digital content, and an example of that is the use of speech recognition and other technologies when creating subtitles for online video, both for entirely automatic subtitling and for further post-editing by a human subtitler. It allows for a faster and more cost-effective subtitling process. It is however important always to remember to put the user in the centre when creating content that is truly accessible for all.

## Författaren

**Michael Stormbom** är affärsområdeschef i det finländska språkföretaget Lingsoft Ab. Han arbetar med affärsutveckling kring lösningar baserade på språkteknologier.



## Källor

- Koponen, Maarit et al, 2020: User perspectives on developing technology-assisted access services in public broadcasting. I: *User perspectives on developing technology-assisted access services in public broadcasting*. Bridge: Trends and Traditions in Translation and Interpreting Studies 1(2), 47-67. <https://www.bridge.ff.ukf.sk/index.php/bridge/article/view/8/25>
- Vázquez et al, 2020: The University of Helsinki Submission to the IWSLT2020 Offline Speech Translation Task. I: *Proceedings of the 17th International Conference on Spoken Language Translation (IWSLT)*, pages 95–102. <https://www.aclweb.org/anthology/2020.iwslt-1.10.pdf>

Europaparlamentets och rådets direktiv (EU) 2016/2102 av den 26 oktober 2016 om tillgänglighet avseende offentliga myndigheters webbplatser och mobila applikationer. Hämtat 23.11.2021. <https://eur-lex.europa.eu/legal-content/SV/TXT/?uri=CELEX:32016L2102>

Hörselskadades riksförbund, Hörselskadade i siffror 2020. Hämtat 23.11.2021. <https://hrf.se/app/uploads/sites/13/2020/12/Statisik-2020-HRF-och-SCB-Horselbron-och-forskning.pdf>

W3C, Web Content Accessibility Guidelines (WCAG) Overview. Hämtat 23.11.2021. <https://www.w3.org/WAI/standards-guidelines/wcag/>

## **Nyckelord**

tillgänglighet, språkteknologi, taligenkänning, textning, AI

# Donera prat för AI-utveckling till Språkbanken i Finland

*Krister Lindén, Mietta Lennes, Tommi Jaubiainen, Mikko Kurimo, Aleksi Rossi*

*Det behövs tusentals timmar med vardagsprat för forskning och innovation i samhällsvetenskap, humaniora och artificiell intelligens. För att utveckla nät-tjänster som förstår svenska och finska i Finland har Språkbanken i Finland tillsammans med Yle, det nationella public service-bolaget för radio, tv och webb, samlat in 4 000 timmar vardagsfinska via nätet. Insamlingen genomfördes med understöd från statens utvecklingsbolag Vake. En insamling av vardagsprat på finlandssvenska är på gång i samarbete med Svenska litteratursällskapet i Finland, och målet är ett liknande projekt även för samiska i Norden.*

## 1. Inledning

Det finns redan kommersiella system som använder AI med taligenkänning, såsom Apples Siri och Googles Alexa. Men många fler idéer väntar på ett kommersiellt genombrott, något som i viss mån beror på bristen på lämpliga språkresurser. En stor, öppet tillgänglig språkresurs möjliggör och påskyndar utvecklingen av språkbaserade AI-komponenter och applikationer. Öppet tillgängliga komponenter sänker tröskeln för att pröva nya idéer, samtidigt som man kan senarelägga beslutet om vilken kommersiell teknik som ska användas i produktionsfasen. Det pågår en världsomspännande insats av Mozilla Common Voice (<https://commonvoice.mozilla.org/>), men det initiativet har som mål att samla in uppläst tal. Färdiga manus tenderar att få människor att använda standardiserat icke-vardagligt tal.

I det projekt vi beskriver var vårt mål specifikt att samla vardagligt, spontant tal från ett stort antal talare. Vi redogör här för den process som ledde till att Vake, det finländska statliga utvecklingsbolaget (numera Ilmatorahasto Oy), beslutade finansiera utvecklingen av en insamlingsplattform för kampanjen *Labjoita puhetta* för insamlingen av finskt talspråk; varför Språkbanken i Finland (<https://www.kielipankki.fi/sprakbanken/>) valdes som dataförmedlare; hur Språkbanken har förberett sig på att göra stora personuppgiftssamlingar tillgängliga och hur det nationella public service-mediebolaget Yle utformade mediekampanjen för att få så många som möjligt att donera prat.

## 2. Applikationer för specialbehov

Att söka i talinspelningar efter innehåll är felbenäget, även om det finns tekniker för att hitta enstaka ord och fraser i talsegment. Ett annat tillvägagångssätt är att konvertera tal till textavskrift och använda befintliga verktyg för textanalys. Man vill kanske till exempel räkna hur många inspelade telefonsamtal som nämner vissa stickord i en robocall-undersökning. Exempel på en mer komplex användning är analyser av samtal för att övervaka telefonförsäljning. En annan tillämpning är automatisk translitterering av intervjuer som journalister eller forskare har gjort. I fall man snabbt kunde hitta ett citat från talsignalen skulle det avsevärt påskynda verifieringen av detaljer i sådana intervjuer. En bättre sökbarhet för talinspelningar ökar också användbarheten av videoinspelade debatter för verifiering i efterhand, exempelvis för debatterna i samband med beslut i riksdagens plenum.

Automatisk taligenkänning används för traditionell diktering, när man behöver skriva meddelanden i situationer där händer och ögon har andra uppgifter. Diktering anpassad till en specifik persons tal fungerar redan någorlunda bra, till exempel på mobila enheter. Det gäller särskilt när bakgrundsljuden är få eller den som talar befinner sig nära mikrofonen.

Med förbättrad talbehandling kan tv-program och föreläsningar textas automatiskt, antingen direkt från originalljudet eller med återdiktering. Specifika grupper, som hörselskadade, skulle ha stor nytta av automatisk textning i realtid. Tillförlitligt fungerande, genreoberoende textning av talspråk är en god grund för automatisk översättning och tolkning, något som har otaliga användningsområden i en globaliserad värld.

Det moderna samhället kräver många digitala användarkompetenser, bland dem färdigheten att använda mobila enheter. Någon med nedsatt syn eller vars fingerfärdighet inte räcker till för en viss enhet, kan därför bli utestängd från många tjänster. Ofta kan kraven som ställs på användarna avhjälpas med ett röstaktiverat användargränssnitt för tjänster på deras modersmål. Intelligent tillämpningar kan komplettera eller till och med ersätta personlig service för åldringar och funktionshindrade och erbjuda dem möjligheter till ett självständigt liv och en förbättrad livskvalitet. Å andra sidan, om det finns ett röstgränssnitt som fungerar dåligt, kan det väcka misstro och leda till att användarna börjar undvika tjänsten. Inom sjukvården kan brister i användargränssnittet rentav utgöra en säkerhetsrisk.

I språkinlärningsapplikationer är talgränssnitt anpassade för specifika användare mer användbara. Interaktion och muntliga färdigheter betonas i dagens samhälle och yrkesliv och blir en allt viktigare del av språkinläringen. Invandrare i Finland har en stor fördel av goda muntliga kunskaper i finska,

både på arbetsmarknaden och för att bygga upp sina sociala nätverk. En stor databas med transkriberat vardagstal är en bra referens. Men det behövs också andra typer av data för att tillförlitligt mäta uttalsegenskaper i enskilda språkinlärares tal och för att modellera deras tal och kommunikation i en verklig interaktion.

Det finns också användningsfall där talet som ska analyseras inte behöver presenteras i textform, utan där analysen härleds direkt från talet. Sådana är till exempel automatisk talaridentifiering eller automatiserad analys av en användares ålder, vakenhetstillstånd eller hälsa. De funktionerna är användbara vid anpassningen av olika applikationer och tjänster, också fastän noggrannheten är lägre än 100 procent. Även när applikationer inte kräver att talet presenteras i textform behövs det ändå stora träningskorpusar med taldata annoterade med personliga och hälsorelaterade egenskaper.

### **3. Behovet av talkorpusar**

I början av 2000-talet var den finländska talteknikens och talspråksforskningens resurser spridda runtom i Finland i relativt små team. USIX – *Uusi käyttäjikeskeinen tietotekniikka* [sv. Det nya användarcentrerade teknikprogrammet för informationsteknik] lanserades 1999 och finansierades av Tekes (numera Business Finland). Programmet resulterade i nya projekt och samarbete mellan forskargrupper och gynnade forskningen inom finländsk tal- och språkteknologi. Med medel från undervisningsministeriet genomfördes en undersökning av talspråksforskningen i Finland (Toivanen och Miettinen 2001). Ett av de viktigaste resultaten var insikten om att det krävdes investeringar i tillgången på digitala taldata för att främja utvecklingen av forskning och teknologi för att kunna bearbeta talad finska.

Tillgången på taldata är en förutsättning för forskning inom talat språk och utveckling av taltekniska applikationer, inklusive talgränssnitt. Syftet med konsortieprojektet *Integrerade resurser för talteknologi och talspråkforskning i Finland* (SA-Puhe), som finansierades av Finlands Akademi 2003–2004, var att ta itu med behovet av allmänna riktlinjer och metoder för forskare att i samverkan samla in, kommentera och dela talkorpus. Under projektet samarbetade fonetiker och språkforskare vid Helsingfors universitet med Laboratoriet för akustik och signalbehandling vid Helsingfors tekniska högskola och med CSC–IT Center for Science.

SA-Puhe-projektet tog tag i behovet av en centraliserad infrastruktur för lagring, delning och underhåll av både taldata och relaterade kommentarer för forskningsändamål. Plattformen skulle byggas på ett objektorienterat databassystem som hade utvecklats vid Tekniska högskolan i Helsingfors inklusive

ytterligare ett samarbete med Helsingfors universitet under 1990-talet (Karjalainen och Altosaar 1993; Altosaar, Millar och Vainio 1999). Databassystemet skulle ha ett grafiskt gränssnitt (Altosaar och Lennes 2005). För att göra det möjligt för forskare att bidra med, dela och underhålla sina utskrifter och strukturerade kommentarer till talinspelningarna, utvecklades en första version av ett talannoteringsprogram (Puh-Editor) vid CSC-IT Center for Science (Grönroos och Miettinen 2004).

Tyvärr var det inte möjligt att slutföra integrationen av komponenterna i taldatabasplattformen och editorn under finansieringsperioden. Under projektets gång togs det dock fram allmänna riktlinjer för talannotering med hjälp av språkforskarna (Lennes och Ahjoniemi 2005). Dessa riktlinjer visade sig vara användbara när idén om big data för talbearbetning återupplivades, inspirerad av framstegen inom talteknologi med neurala nätverksteknologier.

Processen som ledde fram till kampanjen *Lahjoita pubetta* inleddes med möten mellan en ad-hoc-grupp av företag och offentliga organisationer år 2018. Statens utvecklingsbolag Vake beställde våren 2019 en förstudie över behovet av finska språkresurser för artificiell intelligens av FIN-CLARIN och Språkbanken i Finland (Kielipankki). Målet var att specificera vilka insatser som krävdes för att möjliggöra en bred användbarhet av de språk som talas i Finland i olika AI-tillämpningar. Satsningen inleddes med finska, som det mest talade språket i Finland. Språkbanken intervjuade över 50 kommersiella och offentliga organisationer i Finland. Ett av de åtta utvecklingsmål som identifierades var en stor korpus av spontant tal, precis som konstaterats i förstudien publicerad i oktober 2019.

Språkbanken i Finland samarbetade med Finlands nationella mediebolag (Yle) och Statens utvecklingsbolag (Vake Oy) i *Lahjoita pubetta*. Experter från Helsingfors universitet, Aalto-universitetet och Åbo universitet deltog också i projektet. Vake utarbetade dataskyddsanalysen och de juridiska dokumenten med hjälp av företaget 1001 Lakes Oy, och juridiska rådgivare från Helsingfors universitet och Yle bidrog med att utarbeta det juridiska ramverket för insamlingskampanjen.

#### **4. FIN-CLARIN och Språkbanken i Finland**

Sedan 2009 har FIN-CLARIN funnits på den nationella kartan över forskningsinfrastrukturer som upprätthålls av Finlands Akademi. FIN-CLARIN-konsortiet består av alla finska universitet som bedriver språk- och språkteknologisk forskning, Institutet för de inhemska språken (Kotus) samt CSC-IT Centre for Science. FIN-CLARIN upprätthåller Språkbanken i Finland, genom vilken medlemmarna i konsortiet erbjuder olika språkresurser som korpusar, lexikala resurser och verktyg.



Ända sedan Språkbankens tillkomst år 1996 har målet varit att både korpusar och verktyg ska vara tillgängliga för forskarvärlden på ett så effektivt sätt som möjligt. Eftersom det har ägnats liten uppmärksamhet åt att göra material och verktyg tillgängliga för företag, licensieras många språkresurser med en icke-kommersiell begränsning. I många fall har också upphovsrätt eller data-skydd lett till begränsade licenser. I FIN-CLARIN ansvarar CSC för det tekniska underhållet och Helsingfors universitet för anskaffningen och kurateringen av korpus och verktyg.

## **5. Taldata för kommersiella ändamål**

Att transkribera tal till text är en subjektiv process. En utskrift produceras för ett visst syfte och den återspeglar de val som gjorts av en enskild transkribberare. Oavsett transkriptionssystem kan en transkription inte spegla alla relevanta egenskaper för naturlig interaktion och nyanser i talet. Dessa inkluderar tillfälliga variationer i talljud eller andra ljud, liksom grundläggande prosodiska egenskaper: röstkvalitet, tonhöjd, intensitet, talhastighet och pauser. De här egenskaperna bidrar inte bara till intrycken av melodi, accenter och rytm, utan också till de upplevda betydelse, avsikter och attityder vi hör och förstår i varandras tal samt till gester, uttryck, blickar och andra handlingar relaterade till interaktionssituationen och dess sammanhang. Det primära målet för transkriptionen av det insamlade talet är att tillhandahålla en fonematiskt korrekt transkription av ljuden i signalen, som senare konverteras till standardiserat tal för sökbarhet och möjliggör vidare forskning och utveckling i automatiserad språkbearbetning.

Konstruktionen av säkra, persondatavänliga röst användargränssnitt kräver i vissa fall att komponenterna i en applikation kan användas utan överföring av persondata från en tjänst till en annan, till en tredje part eller till en annan stat. Det här talar för att talbearbetningens komponenter bör vara öppet tillgängliga med öppen källkod.

Talkorpusar som distribueras av Språkbanken i Finland, till exempel från Finlands riksdagsplenium, som innehåller inspelningar av riksdagens plenarsammanträden från 2008 till 2020, samt utskrifter av dessa, är licensierade som CC-BY-NC-ND. NC är en förkortning för icke-kommersiell. Den innebär att materialet inte får användas för kommersiella ändamål. Ett sätt att utöka det kommersiellt användbara talmaterialet är att omförhandla licenser för korpusar för att tillåta en sådan användning. Även om det i det nämnda fallet med parlamentets plenarsessioner fortfarande är möjligt, går det ofta inte att omförhandla tillgången till talmaterial efter att det har samlats in. Av den anledningen är det viktigt att se till att nytt talmaterial samlas in på ett sätt som möjliggör kommersiell användning.

## 6. Juridiska aspekter på insamling av talmaterial

Det juridiska ramverket i EU ska tillhandahålla ett interoperabelt utrymme för olika slag av verksamhet. Medan det juridiska ramverket harmoniserar mycket av verksamheten i samhället i övrigt, har forskningspraxis ofta överlåtit till nationella överväganden. Det här påverkar möjligheterna till utbyte av data genom en forskningsinfrastruktur som CLARIN, eftersom vi behöver en gemensam rättslig grund som går att tillämpa på forskning i alla EU-länder. Dessutom är forskningen inte enbart begränsad till den akademiska världen. För att fördela resurserna inom ett land, behövs det därför ofta lösningar som också omfattar industriell och kommersiell forskning.

De immateriella rättigheterna i det juridiska ramverket inom EU har diskuterats ingående (Kelli & al. 2016, 2018b, 2019a) av medlemmar i CLARIN Legal Issues Committee. CLARIN rekommenderar Creative Common-licenser när det är möjligt (Oksanen & al. 2011). För alla datamängder, inklusive dem som inte kan göras öppet tillgängliga, erbjuder CLARIN ett klassificeringssystem för licensmetadata (Oksanen & al. 2010) för att informera användaren om potentiella begränsningar när man får tillgång till en datasamling. För datasamlingar som inte kan göras öppet och offentligt tillgängliga erbjuder CLARIN standardlicensmallar för deponering av data som ska delas via CLARIN (Kelli & al. 2018a). De immateriella rättigheter som är relevanta för att dela forskningsdata har granskats tämligen ingående av CLARIN under de senaste tio åren, men nya möjligheter öppnar sig via EU:s text- och datautvinningsdirektiv (Kelli & al. 2020b).

Under de senaste åren har konsekvenserna av EU:s allmänna dataskyddsförordning (GDPR) blivit allmänt kända (Kelli & al. 2021). Enskilda EU-medlemsländer har ett visst utrymme att göra undantag för forskning, vilket har lett till varierande praxis när det gäller delning av personuppgifter för akademiska forskningsändamål (Kelli & al. 2019b, Lindén & al. 2020). Resurser som innehåller personuppgifter är bland de resurser som inte kan göras tillgängliga utan skyddsåtgärder, och CLARIN uppdaterar nu licensmallar för att återspegla hur personuppgifter fortfarande kan delas säkert och kontrollerat för akademisk forskning (Kelli & al. 2020a). Trots att alla data inte kan göras öppet tillgängliga är det möjligt att använda data som man har laglig tillgång till för att skapa öppet tillgängliga språkmodeller (Kelli & al. 2020c).

Från början stod det klart att behandlingen av personuppgifter i *Labjoita puhetta*-projektet måste bedrivas både juridiskt och etiskt hållbart. Alla centrala aktörer i *Labjoita puhetta*, Språkbanken vid Helsingfors universitet, Vake och Yle är offentliga organisationer som inte kan bortse från sådana aspekter. För att förstå de eventuella problem behandlingen av personuppgifter kunde

medföra för individer gjordes därför en noggrann riskbedömning. Dataskyddsmyndigheten förordar ett balanstest för den som vill yrka på legitimt intresse som laglig grund för insamling. Efter att ha gått igenom de sex stegen i testet, stod det klart att insamlingen uppfyllde kraven och att intressena i forskningen inte åsidosattes vare sig av donatorernas intressen eller deras grundläggande rättigheter och friheter. Det gjordes också en konsekvensbedömning av dataskyddet med tanke på möjliga risker i anslutning till behandlingen av personuppgifter, där särskilt den omfattande bearbetningen av materialet och utvecklandet av ny teknologi relaterades till ändamålet. Språkbanken i Finland kommer att ge tillgång till taldata när en tillräcklig mängd material har donerats och processen med att ge rättigheter till kommersiella aktörer är slutförd. Detta borde vara möjligt i början av år 2022.

En närmare beskrivning av de juridiska dokument som utarbetades för insamlingen av personuppgifter så att taldata kan delas inom EU, finns i Lindén & al. (2022). Det faktum att uppgifterna samlades in också för industrins bruk gör att avtalsramverket är relevant även i nordiska sammanhang, eftersom den industriella användningen regleras av EU:s gemensamma dataskydd.

## **7. Hur man planerar en massiv taldatainsamling**

Den första fasen av *Lahjoita puhetta*-kampanjen fokuserade på att få ihop 10 000 timmar vardagsfinska representativ för hur finländarna i dag talar i olika sammanhang.

Kampanjen hade en föregångare i ny metodik och nya sätt att erhålla taldata över internet som hette *Prosovar* och genomfördes av forskare vid Åbo universitet. Målet med *Lahjoita puhetta* var inte bara att samla in en stor mängd vardagstal, utan även att nå ut till så många som möjligt. I marknadsföringen betonades det att donationer av alla varianter av talad finska var välkomna, också från talare som inte hade finska som modersmål. Dataskyddet och instruktionerna förutsatte ändå en viss nivå på språkkunskaperna.

För att hitta en balans mellan de materiella målen, de tekniska möjligheterna och de tillgängliga resurserna hölls designworkshoppar för alla intresserade. Under dem samlade man in idéer för både insamlingskoncept och användningsområden från forskare både inom industri- och universitetsvärlden. Största delen av den praktiska planeringen överläts till personalen på Yle, som ansvarade för att mobilisera allmänheten genom sina radio- och tv-kanaler. Yle designade bilderna, filmerna och texterna i webbapplikationen och de nedladdningsbara apparna. Det utformades också tekniska mallar för design av teman med olika typer av innehåll riktade till specifika målgrupper.

Hösten 2019 och våren 2020 bestämdes potentiella användningar, målgrupper samt obligatoriska och valfria funktioner i gränssnittet på workshoppar, som organiserades av utvecklingsföretaget Solita enligt en metodik som delvis byggde på Design Thinking. Senare testades en rad nyckelfunktioner med pappersprototyper och semi-interaktiva verktyg. Förslag gjordes av professionella och erfarna webbservicedesigners, och en del funktioner testades i praktiken.

Nyckelfrågorna och utmaningarna för utformningen av användargränssnittet gällde att få personerna att tala fritt och vinna deras förtroende så att de kände sig bekväma, samtidigt som gränssnittet måste uppfylla de juridiska begränsningarna och enkelt presentera den nödvändiga informationen. Efter att ha testat några idéer bestämde sig Yle för en video, en bild och ett textinnehåll som lockade en person att prata, kombinerade med en funktionell knapp med vilken talaren kunde starta och stoppa inspelningen.

Det fördes också diskussioner kring något slag av spelelement som meddelade användarna hur mycket de hade donerat, till exempel i form av jämförbara resultat för att upprätthålla intresset eller sociala element som att dela resultat eller samla team. Slutligen inkluderades enbart den totala tid som donerats som ett sådant spelelement.

Inledningstemat *Harjoitellaan ensin* (sv. Vi börjar med att öva) startade med en testsession av inspelningen med användaren. Sessionen var samtidigt ett utmärkt tillfälle att betona att främst AI-forskare skulle använda inspelningarna och att påminna användaren om dataskyddet. Den tekniska plattformen presenterade metadatafrågor om dialektbakgrund och grundläggande demografi som åldersgrupp, kön, modersmål, nuvarande boningsort, yrkeskategori och utbildningsnivå. Även användarens tekniska plattform noterades för statistiska ändamål, men inga positionsdata samlades in.

Yle tog slutligen fram ett 40-tal okomplicerade teman för att stimulera insamlingen. Utöver övningstemat samlades nästan hälften av uppgifterna in med följande av de tolv populäraste temarubrikerna:

Rakkain eläimeni (*Mitt käraste djur*)

Mistä kodikkuus syntyy? (*Vad gör ett hem trevligt?*)

Tärkeä esineeni (*Ett viktigt/kärt ting*)

Lempivaate (*Mitt favoritplagg*)

Mikä suututtaa? (*Vad gör mig arg?*)

Turhat tavarani (*Mina onödiga prylar*)

Mitä opimme? (*Vad lärde vi oss?*)

Entisajan lemmikit (*Husdjur från förr*)

Katson ikkunasta (*Vad jag ser från mitt fönster*)

Kuva-arvoitus (*En bildgåta*)

Kerro aamiaisesta (*Berätta om din frukost*)



*Bild 1. Mitt käraste djur var ett populärt tema i pratdonationerna. Foto: Språkbanken*

Yle gjorde humoristiska inforeklamer med uppmaningar till allmänheten om att donera tal. Dessa sändes mellan programmen i radio- och tv-kanalerna sommaren och hösten 2020, med några repriser våren 2021.

## **8. Tekniskt genomförande**

Talet till *Lahjoita puhetta* (<https://lahjoitapuhetta.fi/>) gick att donera via webbläsare eller mobilappar. Representanter både från industri och universitetsvärlden utvecklade de allmänna specifikationerna. Företaget Solita utvecklade apparna. Mjukvaruplattformen finns publicerad som en öppen källkod som tillåter andra organisationer att bygga egna system för talinsamlingar, specialiserade insamlingskampanjer eller för liknande kampanjer i andra länder.

Teknisk röstkvalitet är ett komplicerat ämne. Det är till exempel nödvändigt att mikrofonen finns nära användaren. Goda råd med avsikten att få talarna att slappna av genom att låta telefonen ligga på bordet hade lett till mer eko och en svagare signal. Att låta en grupp människor diskutera var också uteslutet. En fritt flödande gruppdiskussion hade visserligen haft uppenbara fördelar jämfört med en talsituation med endast en talare. Den hade däremot inneburit svåra tekniska utmaningar både med deltagarna nära varandra, med samma mikrofon, och långt borta från varandra med var sin enhet för att minimera störningar och eko. Dessutom hade detta krävt en synkronisering av flera signaler eller

en registrering av vilka telefoner som spelade in diskussionen. Av den här anledningen genomfördes inga tester med gränssnitt och teman för dialogflöden.

Inspelningarna hölls enkla, genom att bara spela in talsignalen i högsta möjliga förlustfria format och lagra den med metadata om system, telefonmodell och version. Metadata möjliggör vissa efterbearbetningskorrigeringar med till exempel ljudutjämning per mikrofontyp. En rudimentär mätare för att ge feedback till användaren om en acceptabel signalnivå övervägdes, men den implementerades inte, bland annat för att spara på mobilernas batterier. Användartester gav oss också orsak att betvivla ändamålsenligheten av en sådan mätare. För det första skulle mätaren ha utgjort en distraktion eller med största sannolikhet ha ignorerats. För det andra skulle informationen om hur mätaren skulle tolkas ha belastat användargränssnittet och för det tredje skulle förbättringen av signalen inte ha varit väsentlig, eftersom användaren förmodligen bara hade talat närmare mikrofonen under en kort stund.

Till slut instruerades användarna helt enkelt att tala fritt i sin egen miljö. En tydlig signal i en störningsfri miljö är ofta att föredra. För tillfället har inspelningarna lite mer variation när de innehåller en del störningar, som bakgrundsljud från andra människor eller vind utomhus. Enligt användartesterna gjordes de flesta inspelningssessionerna i tysta inomhusmiljöer. En fördröjd sändning i bakgrunden av lokalt lagrade inspelningar för uppladdning till molnet förbereddes ifall användaren inte hade en stabil internetuppkoppling, men det var förmodligen inte en avgörande funktion.

Webben, Android och iOS valdes som plattformar för smartphones, surfplattor och datorer med mikrofoner. Den tillhörande webbplatsen informerade också användarna om kampanjen och Yle publicerade artiklar på sin kampanjsajt. För att väcka användarnas förtroende för kampanjen gavs apparna ut via Yle istället för via separata dedikerade eller kampanjspecifika app-konton.

Servicearkitekturen bestod av flera gränssnitt på olika plattformar, webbservicer och databaser för insamling av data i molnet. Genom att fördela ansvaret mellan Yle och Helsingfors universitet kunde det juridiska ansvaret för de olika parterna begränsas, så att Yle som genomförde kampanjen fick tillgång till aggregerade användningsdata utan att få tillgång till och ansvar för persondata. Systemet utvecklades för enspråkigt bruk. Anpassningen och lokaliseringen till andra språk som finlandssvenska och samiska hölls i åtanke och testas under 2022.

För att beakta GDPR och möjliggöra radering av bidrag tilläts enkel radering av användarbidrag genom en lång slumpmässig identifierare som ges till användaren vid taldonationen. Det finns inga andra användarspecifika identifierare. Man måste ändå alltid överväga om enskilda användare kan identifieras med sina metadata ifall den deltagande gruppen är liten eller en kombination av vissa

metadata är mycket specifik. Exempelvis kan antalet män i en viss åldersgrupp på ett litet geografiskt område med en speciell dialektbakgrund potentiellt resultera i att en grupp människor skulle kunna identifieras både med den insamlade informationen och i den verkliga världen. Den tekniska plattformen som sådan begränsar inte insamlingen av specifika metadata eftersom GDPR-kompatibel behandling av personuppgifterna sker på ansvar av den som senare behandlar uppgifterna eller publicerar resultat som bygger på dessa.

Under våren 2021 deltog kampanjens Android- och iOS-mobilapplikationer i den årliga marknadsföringstävlingen *Grand One* för webbapplikationer. Applikationerna vann första pris i kategorin mobiltjänster och fick ett hedersomnämmande i kategorin för bästa dataanvändning. Yle skickade också in kampanjen till den årliga tävlingen för europeiska mediebolag *Prix Europa*, där kampanjen vann kategorin *Bästa europeiska digitala ljudprojekt 2021* för tv-, radio- och online-produkter bland 684 bidrag från 26 länder.

## 9. Det insamlade materialet

Kampanjen *Lahjoita puhetta* engagerade över 25 000 medborgare i Finland. De donerade över 240 000 talprover med totalt cirka 4 000 timmar vardagstal som kan användas både akademiskt och industriellt för att utveckla och forska i språk- och AI-applikationer. Cirka 3 500 timmar samlades in på bara ett halvår och ytterligare 500 timmar under halvåret därpå. Lanseringen på nationell tv i juni 2020 inspirerade till det största antalet bidrag, med närmare 500 timmar under de första månaderna. Kampanjen nådde nya målgrupper under hösten 2020, om än i långsammare takt. Mot slutet av kampanjen fokuserade man på att samla in dialekter med hjälp av regionalradion och ytterligare tio procent av materialet samlades in under en vecka kring julen 2020. Yle inrättade en kampanjsida för sina kampanjevenemang (<https://yle.fi/aihe/lahjoita-puhetta>). Kampanjen avslutades officiellt vid nyår 2020 men infoblänkare och repriserna sändes fortfarande under våren 2021, vilket resulterade i ytterligare en rännil av bidrag. Närmare 90 procent av bidragen var mellan 10 sekunder och 3 minuter, med en medianlängd per inspelning på 30–60 sekunder.

Det samlades taldata från ett brett spektrum av åldersgrupper. Lite överraskande var 21–30-åriga kvinnor oberörda av det något tekniska upplägget och donerade det mesta av talet. Den minsta mängden tal donerades av mycket unga (1–10 år gamla) och mycket gamla (80 år eller äldre). Alla har inte tillhandahållit alla metadata, men bland dem som uppgett metadata fanns några intressanta observationer: Personer mellan 20 och 60 år gjorde cirka tre fjärdedelar av donationerna. Nästan 70 procent av donatorerna var kvinnor. Som väntat kom närmare hälften av donationerna från fyra regioner med de största städerna: Nyland

(inklusive Helsingfors och Esbo), Norra Österbotten (inklusive Uleåborg), Egentliga Finland (inklusive Åbo) och Birkaland (inklusive Tammerfors). Donationer gjordes ändå i alla regioner i Finland i över 50 städer. 95 procent av donatorerna var infödda finländare. De geografiska områdena hade ungefär samma mängd donationer per 100 000 invånare. En stor andel invånare med svenska och samiska som modersmål i vissa områden förklarar troligen ett par regioner med lite färre bidrag. Mer än två tredjedelar av uppgifterna donerades av studerande, pensionärer, lärare, entreprenörer, experter och sjukskötare (i fallande ordning efter bidragsgivarnas antal) och resten kom från mer än 30 andra yrkesgrupper. Cirka 62 procent hade högre utbildning och 28 procent gymnasieutbildning.

## 10. Avslutningsvis

Trots ett ambitiöst mål lyckades kampanjen *Labjoita pubetta* samla in en omfattande resurs av finskt talspråk från ett stort antal talare på bara några månader. Donationer kom från alla regioner i Finland och de geografiska områdena hade ungefär samma mängd donationer per 100 000 invånare. Av de insamlade 4000 timmarna har 1 500 timmar translittererats.

Målet med kampanjen var inte bara att samla in en stor mängd tal, utan att nå ut till så många olika grupper och till så många individer som möjligt. Alla donatorer tillhandahöll inte fullständiga metadata, men bland dem som gav metadata noteras att personer mellan 20 och 60 år utgjorde cirka tre fjärdedelar av donationerna och att nästan 70 procent var kvinnor. Två grupper att överväga för framtida punktinsatser är tonåringar runt 11–20 och pensionärer runt 71–80. Bägge har distinkta egenskaper ur AI-utvecklingssynpunkt, talar i olika tonhöjd, med olika ordförråd, takt, pauser och andning. Företag överväger att utveckla AI-drivna äldreomsorgssystem, och specifika taldata för personer som är sängliggande kunde därför också vara användbara.

En liknande kampanj *Donera prat* för att samla in finlandssvenskt vardagspråk med samma syfte lanserades av Yle i slutet av 2021 i samarbete med Helsingfors universitet och Svenska litteratursällskapet i Finland. Målet är att förbättra möjligheterna att utveckla talbaserade tjänster som förstår finlandssvenska.

En insamling av samiska planeras i samarbete med Yle, Norges public service-radio och -television NRK och universiteten i Helsingfors och Tromsø.

## Summary

Thousands of hours of everyday talk are needed for research and innovation in social sciences, humanities and artificial intelligence. To develop online services that understand everyday Finnish, the Language Bank of Finland together with the



national broadcasting company Yle has collected 4,000 hours of colloquial Finnish online with funding from the Finnish State Development Company Vake to develop the speech collection platform. A similar campaign is underway for Finnish-Swedish in collaboration with the The Society of Swedish Literature in Finland (SLS) and the goal is also to carry out a similar campaign for the Sami languages in the Nordic countries in cross-border cooperation with interested parties.

## Författare

Forskningsdirektör **Krister Lindén** är nationell koordinator för FIN-CLARIN och verksamhetsledare för Språkbanken i Finland. Dr **Tommi Jauhainen** och MA **Mietta Lennes** är projektplanerare på Språkbanken i Finland vid Helsingfors universitet. Prof. **Mikko Kurimo** är professor i talteknologi vid Aalto universitet. MSc **Aleksi Rossi** är utvecklingsdirektör på det nationella mediebolaget Yle.



*Krister Lindén*



*Tommi Jauhainen*



*Mietta Lennes*



*Mikko Kurimo*



*Aleksi Rossi*

## Källhänvisningar

- Altosaar, T., B. Millar & M. Vainio, 1999: Relational vs. object-oriented models for representing speech: a comparison using ANDOSL data. I: Proceedings of EUROSPEECH'99, Budapest, Hungary, 5-9 Sep 1991, Vol. 2, pp. 915-918.
- Altosaar, T. & M. Lennes, 2005: A Graphical Query Formation Compiler for Speech Database Access. I: The Second Baltic Conference on Human Language Technologies, Tallinn, Estonia, April 4-5, 2005, pp. 209-218.
- Grönroos, M. & M. Miettinen, 2004: Infrastructure for Collaborative Annotation of Speech. I: LREC 2004, pp. 543-546.
- Karjalainen, M., & T. Altosaar, 1993: An Object-Oriented Database for Speech Processing. I: Proceedings of Eurospeech 1993, Madrid, Spain.
- Kelli, A., K. Vider & K. Lindén, 2016: The regulatory and contractual framework as an integral part of the CLARIN infrastructure.
- Kelli, A., K. Lindén, K. Vider, P. Labropoulou, E. Ketzan, P. Kamocki & P. Straňák, 2018a: Implementation of an Open Science Policy in the context of management of CLARIN language resources: a need for changes? I: Selected papers from the CLARIN Annual Conference 2017, Budapest, 18-20 September 2017. Linköping University Electronic Press.
- Kelli, A., T. Mets, K. Vider, A. Värvi, L. Jonsson, K. Lindén & R. Birštonas, 2018b: Challenges of transformation of research data into open data: The perspective of social sciences and humanities. *International Journal of Technology Management & Sustainable Development*, 17(3), 227-251.
- Kelli, A., A. Tavast, K. Lindén, K. Vider, R. Birštonas, P. Labropoulou, I. Kull, G. Tavits & A. Värvi, 2019a: The extent of legal control over language data: the case of language technologies. I: Proceedings of CLARIN Annual Conference 2019. CLARIN ERIC.
- Kelli, A., K. Lindén, K. Vider, P. Kamocki, R. Birštonas, S. Calamai, P. Labropoulou, M. Gavrilidou & P. Straňák, 2019b: Processing personal data without the consent of the data subject for the development and use of language resources. I: Selected papers from the CLARIN annual conference 2018, Pisa, 8-10 October 2018 (pp. 72-82). Linköping University Electronic Press.
- Kelli, A., K. Lindén, K. Vider, P. Kamocki, A. Tavast, R. Birštonas, G. Tavits, M. Keskküla & P. Labropoulou, 2020a: CLARIN contractual framework for sharing language data: the perspective of personal data protection. I: Pro-

ceedings of CLARIN Annual Conference 2020. 05–07 October 2020, Online Edition. CLARIN ERIC.

Kelli, A., A. Tavast, K. Lindén, K. Vider, R. Birštonas, P. Labropoulou, I. Kull, G. Tavits, A. Värvi, P. Straňák & J. Hajic, 2020b: The Impact of Copyright and Personal Data Laws on the Creation and Use of Models for Language Technologies. I: Selected Papers from the CLARIN Annual Conference 2019. Linköping University Electronic Press.

Kelli, A., A. Tavast, K. Lindén, R. Bristonas, P. Labropoulou, K. Vider, I. Kull, G. Tavits, A. Värvi & V. Mantrov, 2020c: Impact of Legal Status of Data on Development of Data-Intensive Products: Example of Language Technologies. Legal Science: Functions, Significance and Future in Legal Systems II.

Kelli, A., K. Lindén, K. Vider, P. Kamocki, A. Tavast, R. Birštonas, G. Tavits, M. Keskküla, P. Labropoulou, I. Kull, A. Värvi, M. Erikson, A. Vutt & S. Calamai, 2021: Sharing is caring: a legal perspective on sharing language data containing personal data and the division of liability between researchers and research organisations. I: Selected Papers from the CLARIN Annual Conference 2020. Virtual Event, 2020, 5-7 October (pp. 129-147). Linköping University Electronic Press.

Lennes, M. & S. Ahjoniemi, 2005: Puheaineiston annotaatio eli nimikointi (Version 1.01) [Annotating speech data (Version 1.0)]. Zenodo. <http://doi.org/10.5281/zenodo.1205453>

Lindén, K., T. Jauhiainen, M. Lennes, M. Kurimo, A. Rossi, T. Kurki & O. Pitkänen, 2022: Donate Speech – Collecting and sharing a large-scale speech database for Social Sciences, Humanities and Artificial Intelligence research and innovation. I: A. Witt & D. Fisher (eds.): The CLARIN Book. Berlin: De Gruyter.

Lindén, K., A. Kelli & A. Nousias, 2020: A CLARIN Contractual Framework for Sharing Personal Data for Scientific Research. I: Selected Papers from the CLARIN Annual Conference 2019. Linköping University Electronic Press.

Oksanen, V., K. Lindén & H. Westerlund, 2010: Laundry Symbols and License Management: Practical Considerations for the Distribution of LRs based on experiences from CLARIN. I: Proceedings of LREC 2010 Workshop on Language Resources: From Storyboard to Sustainability and LR Lifecycle Management.

Oksanen, V. & K. Lindén, 2011: Open Content Licenses: How to choose the right one. NEALT Proceedings Series.

Toivanen, J. & M. Miettinen, 2001: Puheentutkimuksen resurssit Suomessa [Resources for speech research in Finland]. CSC – IT Centre for Science Ltd., 2001. ISBN 952-9821-76-X

## **Nyckelord**

infrastruktur, språkbank, korpus, talspråk, massiva datasamlingar

# Språk(teknologi) är nyckeln till (artificiell) intelligens och rättvisa

Jörg Tiedemann

*Vårt samhälle är fullt av ogjorda översättningsjobb. Utan teknologiska lösningar finns det helt enkelt ingen möjlighet att kunna hantera den gigantiska mängd information som borde göras tillgänglig för alla. I artikeln presenteras initiativ som bygger på öppna och säkra översättningsverktyg som beaktar användarnas integritet. Vårt samhälle måste stödja sådana, för att skapa rättvisa utan exploatering, och för att erbjuda tjänster på lika villkor även till mindre kommersiellt intressanta minoriteter.*

Språk spelar en väsentlig roll i vårt mänskliga liv. Jag tvekar inte att påstå att språk är nyckeln till det vi kallar intelligens. Vårt kollektiva kunnande bygger på den omfattande möjligheten att kommunicera och dokumentera allt vi kan. Utan språk skulle vi människor inte ens komma i närheten av de färdigheter som ledde till vår dominans på jorden. Vi använder språk för att beskriva, analysera, debattera och utreda alla världens problem. Vi använder språk för att begripa, lära ut och studera nya och gamla fenomen. Det vi skrivit ner eller berättat vidare sedan generationer tillbaka används för att utveckla vårt samhälle så att vi snabbt kan anpassa oss till nya omständigheter och förändringar.

Språk fyller även en avgörande social roll. Via vårt sätt att kommunicera skapar vi gemenskap, sociala förhållanden och ett nätverk som gör oss starkare än individer var för sig. Men språk kan förstås också missbrukas, det kan manipulera våra tankar och leda till mycket missförstånd. Språk används också för att avgränsa oss från varandra och kan leda till stor irritation. Språk är med oss hela tiden och bestämmer vår vardag hela livet ut.

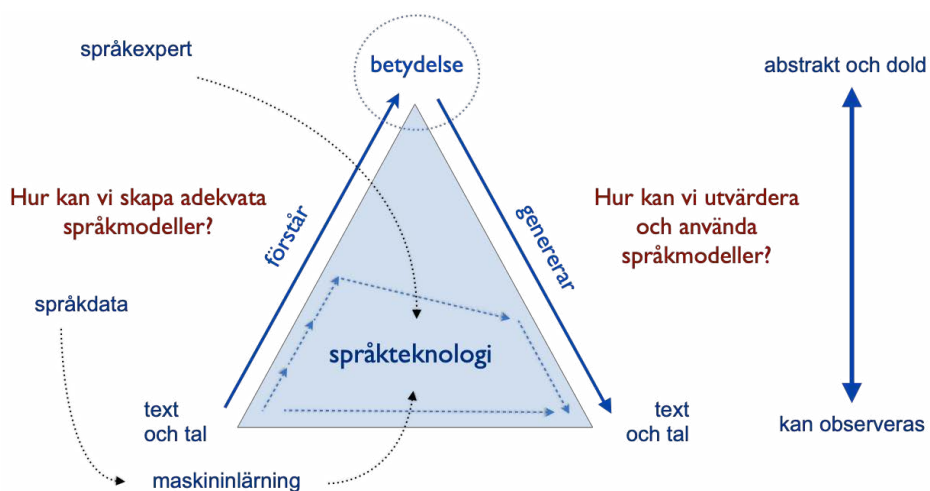
När man betraktar språkens roll är det självklart att vi måste värna om våra språk och hur vi kommunicerar med varandra. Det borde vara självklart att vårt samhälle behöver investera omfattande resurser på språkhantering och effektiv kommunikation. Information måste nå ut till alla för att skapa jämlikhet och för att undvika klyftor, som till stor del bottnar i missförstånd och språkrelaterad diskriminering. Jag vill argumentera för den essentiella roll som öppna språkteknologiska lösningar har i vårt samhälle, där kommunikation sker alltmer globalt och via digitala medier. Och inte bara där, utan lokalt och

inom alla kretsar finns det en utmaning att förmedla information, förhindra misskommunikation och undvika spänningar mellan människor som missförstår varandra. Språkteknologi styr redan nu mycket av vår vardag även om vi inte alltid är så medvetna om det. Det är viktigt att vi kanaliserar denna teknologi till samhällsnytta och rättvisa i stället för exploatering och manipulation.

Jag tar i den här artikeln upp exemplet med maskinöversättning och dess funktion med fokus på informationstillgång och integration och berättar om vårt initiativ att öppna automatiska översättningstjänster så att denna teknologi kan tjäna alla, även små minoriteter som inte finns på radarn för kommersiella produkter. Det finns fortfarande en hel del utmaningar och problem, men de stora fördelarna är tydliga redan nu.

## 1. Var står vi inom språkteknologin just nu?

Den stora utmaningen inom språkteknologin är att skapa datamodeller som förstår och genererar språk som människor gör. Syftet är att kunna interagera språkligt med en maskin på ett naturligt sätt, men också att förstå hur mänskligt språk fungerar som kommunikationsmedel och hur det relateras till betydelsen och kunskapen vi vill förmedla. Det finns två möjligheter att skapa sådana modeller: (1) En expert lägger till allt vi vet om språk och kommunikation i ett dataprogram så att det kan köras för att hantera text och tal. (2) Vi bygger en maskin som kan lära sig att hantera språk med hjälp av data och exempel och denna maskininlärningsmodell suger in allt den kan hitta i text och tal och kontexten som kan observeras. Båda metoderna används flitigt och kombineras även på många olika sätt.



Figur 1. Målsättningar och metoder inom språkteknologi

En trend som har lett till stora genombrott är maskininlärningsmetoder som bygger på artificiella neurala nätverk (ANN). Allt större nätverk tränas i dag med hjälp av beräkningsmaskiner som blir kraftigare varje år, och den växande datamängden gör det möjligt att inkludera mer och mer av vår kunskap som vi sparar i text och inspelat tal.

Den generella principen är inte särskilt märkvärdig: Ett neuralt nätverk kan betraktas som en universell metod för att approximera vilken matematisk funktion som helst, och det är det vi behöver inom den digitala värld vi jobbar med. En dator behöver en sådan funktion som mappar språkliga signaler till numeriska värden – det är det enda som den kan processa. Till exempel blir all text vi knappar in koderad enligt ett fast schema och bokstäverna vi ser på skärmen avkodar denna funktion för att visa en bild som motsvarar våra förväntningar.

Tyvärr finns det ingen exakt funktion som definierar hur språk mappas till betydelse och där kommer inlärningsmetoden in. Även människor lär sig språk genom att använda det och relaterar dess användning till kontexten. Precis så gör också neurala nätverk och de börjar att lära sig användningen av språkliga tecken genom att jämföra med exempel i olika kontexter. Det krävs mycket data för att förstå hur olika koncept ska användas. Dagens neurala språkmodeller matas med miljontals meningar och dokument för att bygga den komplexa funktion som krävs. Internt har denna funktion miljoner eller miljarder okända parametrar som måste optimeras och det krävs en konkret uppgift som bestämmer i vilken riktning optimeringen ska ske.

Det finns olika uppgifter vi kan definiera. Oftast är de väldigt primitiva, som till exempel lucktextövningar där vår maskin måste gissa vilka ord som saknas inom kontexten. Sådana enkla övningar kan enkelt genereras genom att slumpmässigt ta bort ord ur existerande texter. Sedan körs övningen tills systemet inte längre lär sig någonting. Man förvånas över hur mycket ett system kan lära sig om språk med sådana enkla uppgifter. Det är nämligen det som ligger bakom de flesta stora "AI-lösningar" som omtalas flitigt inom medierna, så som BERT, GPT-3 med flera.

Efter all baskunskap sådana modeller har lärt sig kan vi bygga vidare och träna dem i andra uppgifter med relativt lite träningsmaterial – och plötsligt kan våra artificiella språkmodeller resonera om språkliga frågor, hämta information och generera text som passar in i en kontext. Det är imponerande färdigheter som lätt kan blanda oss och få oss att se en allmän intelligens inom systemet som egentligen inte finns där.

Neurala nätverk vet inte vad de gör och varför. Det finns ingen motivation, ingen taktik och inga väl övervägda tankar bakom det som nätet spottar ur

sig. Det reflekterar kunskapen vi lägger in i träningsmaterialet och kan generalisera den på ett effektivt sätt så att det dyker upp mönster och indirekta relationer ur materialet som vi inte förväntade oss att sådana primitiva system skulle kunna hitta.

Datadrivna metoder och maskininlärning med neurala nätverk har lett till en revolution som möjliggör tillämpningar vi bara kunde drömma om för några år sedan. Tal och text kan nu hanteras på ett sätt som är flytande och naturligt. Men vi måste inse att vi fortfarande talar om en avancerad signalprocessering och *inte* om intelligens, i alla fall inte på samma sätt som vi betraktar mänsklig intelligens. Om man ser på utvecklingen av språkmodellering kan vi jämföra den med tre olika nivåer av språkförståelse:

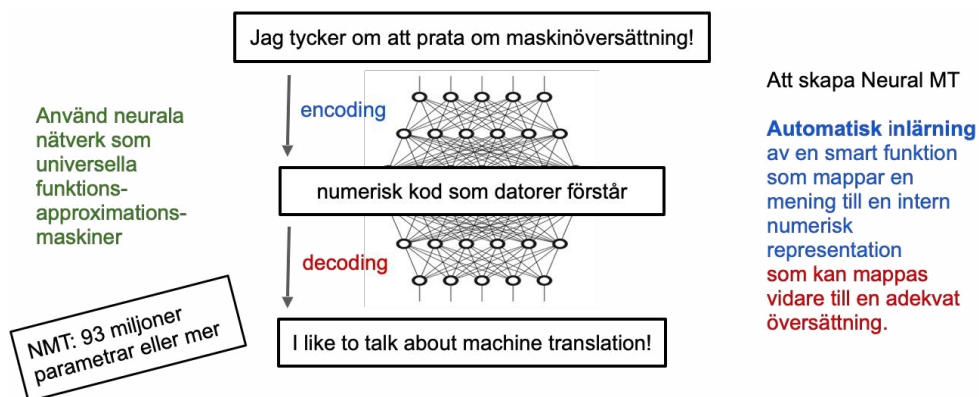
1. Jag hör dig
2. Jag förstår ditt språk
3. Jag begriper vad du säger

Nivå två syftar på kunskapen om språkliga uttryck, hur de används och relateras till varandra. Nivå tre däremot kräver en bild av världen och en modell av intention och syfte med kommunikationen. Språkteknologin har nått steg två för resursrika språk, men har fortfarande en hel del kvar för att uppnå steg tre och att täcka många språk som inte ens får stöd på grundnivån. Det är det vi jobbar med.

## **2. Hur fungerar neural maskinöversättning (NMT)?**

Maskinöversättning kombinerar språkförståelse med språkgenerering. Den är därför en spännande tillämpning som återspeglar de flesta problem vi måste åtgärda inom språkteknologi. Dessutom översätts det kontinuerligt och vi kan lätt samla in exempeldata som hela tiden produceras av mer eller mindre professionella översättare. På samma sätt som jag beskrev de neurala nätverken behöver vi lära oss en komplex matematisk funktion som mappar språk till en numerisk representation. Den kan sedan användas för att generera text (eller tal), men denna gång på ett annat språk. Modeller som kombinerar en kodningsfunktion och en avkodningsfunktion för språkliga data kallas sekvens-till-sekvens-modeller (förkortas seq2seq) och tränas i princip på samma sätt som andra neurala nätverk – men nu med hjälp av en annan träningsuppgift, nämligen översättning.





Figur 2. Schematisk illustration av neural maskinöversättning

Så i stället för lucktexter övar vår seq2seq-modell att generera översättningar som systemet sedan jämför med det en människa har gjort för att anpassa sina interna parametrar, ända tills det inte längre sker någon förbättring. Återigen kan ett sådant system generalisera och hitta mönster som gör det möjligt att översätta mer än bara allt det som finns i träningsmaterialet. Liksom en människa kan hitta nya sätt att kombinera språkliga element så gör vårt översättningsmodell detsamma. Det hela drivs av statistik, frekvens och kontextuella analogier eller så kallad “distributionell semantik”. Den aktuella standardmodellen kallas “transformer” och bygger huvudsakligen på en funktion som kallas för “attention”, där systemet lär sig att betrakta olika delar av den språkliga kontexten för att bestämma vad det ska göra som nästa steg (Vaswani et. al, 2017). Miljoner eller miljarder av parametrar styr översättningen och miljontals exempelmeningar med översättning krävs för att få vettiga översättningsmodeller.

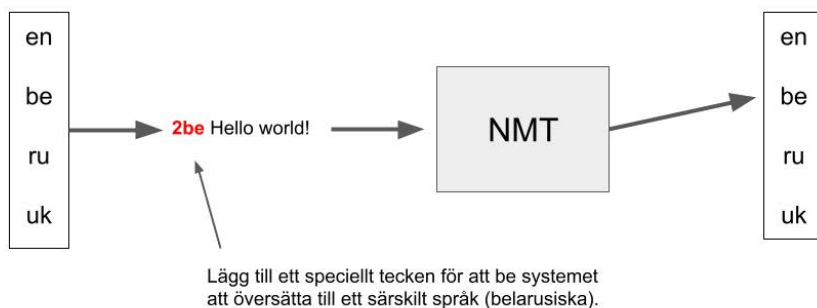
Själva modellen tränas i många iterationer och blir stegvis allt bättre med hänsyn till träningsdatan den använder. Som med alla andra språkmodeller kan vi alltid fortsätta träna när vi får nya data eller behöver förbättra systemet. På det sättet kan ett sådant system anpassas till nya tillämpningsområden. Det är också möjligt att träna på många språk samtidigt. Vi kan skapa en multikulturell polyglott som förstår sig på många språk och kan översätta till många andra språk. Det spännande med denna idé är att flerspråkiga system kan utnyttja likheter mellan olika språk för att överföra kunskap och information från ett språk till ett annat. En efterföljande finjustering kan sedan hjälpa att anpassa modellen till speciella uppgifter och språk. En förutsättning för att det hela fungerar är att vi har tillräckligt med träningsmaterial: Data är därför flaskhalsen.

### 3. Neural maskinöversättning med lite data

Fungerande översättningsmodeller finns det fortfarande bara för ett fåtal av alla världens språk. Inlärningen är krävande och det behövs mycket träningsmaterial. Vi har ett projekt på gång, OPUS<sup>1</sup> (Tiedemann, 2016), som ska förbättra situationen för många språk genom att systematiskt samla in och förbereda material som vi kan dela med oss till alla som jobbar med maskinöversättning, språkteknologi och allmän språkvetenskap. OPUS växer för varje år, men vi måste ändå hitta andra lösningar för språk där vi aldrig kommer att uppnå de resurser som krävs för de standardlösningar som nämndes tidigare. Mycket forskning pågår och det finns åtminstone två populära strategier:

1. Flerspråkiga modeller och kunskapsöverföring
2. Generering av artificiella data

Att skapa flerspråkiga modeller är ganska enkelt. På källspråkssidan kan man enkelt blanda alla språk och det neurala nätet lär sig självt att känna igen olika språk. För att generera rätt målspråk i översättningen kan man använda ett enkelt trick genom att lägga till ett särskilt tecken som berättar för systemet vilket språk man vill att det ska generera (Johnson et. al, 2017). Det här fungerar som när man först talar om för en mänsklig översättare vilket språk man vill hen ska översätta till, och det fungerar förvånansvärt bra även för neurala översättningssystem. Följande bild illustrerar upplägget:



Figur 3. Flerspråkiga översättningsmodeller

Grundtanken med flerspråkiga modeller är att vi slipper träna många olika system för varje översättningsriktning och att likheterna mellan språk kan hjälpa till att förbättra kvaliteten. Illustrationen visar ett exempel med östslaviska

1 <https://opus.nlpl.eu/>

språk och engelska. Språk med mindre resurser, som belarusiska (be), kan dra stor nytta av likheterna med ryskan och ukrainskan när vi tränar översättning till och från engelska.

Denna effekt kan åskådliggöras i ett test vi gjorde med olika tvåspråkiga och flerspråkiga system. Följande tabell ger automatiska utvärderingsvärden för översättningen mellan belarusiska och engelska för en testkorpus med 2 500 meningar (Tiedemann, 2020). Vi använder BLEU (Papineni, 2002) som mått för att jämföra automatiska översättningar med mänskliga referensöversättningar och högre siffror betyder bättre kvalitet. Tabellen visar resultat från modeller med ett stigande antal källspråk som inkluderas i träningsprocessen.

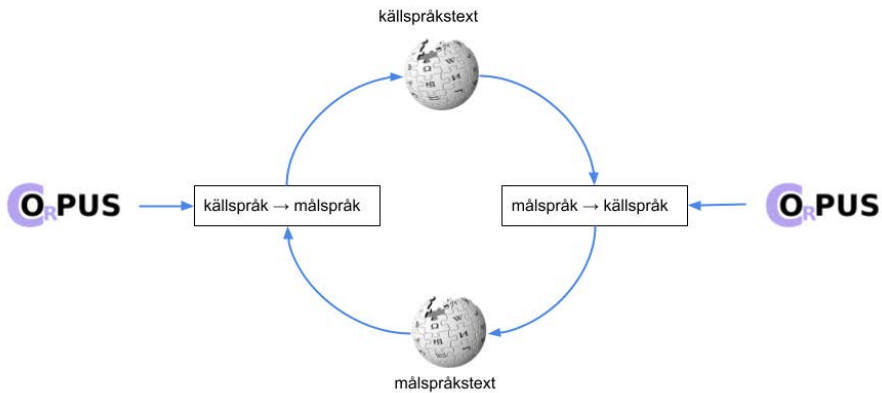
Tabell 1. Automatisk utvärdering av flerspråkiga översättningsmodeller (BLEU värden i %)

modell	belarusiska → engelska	engelska → belarusiska
belarusiska – engelska	10.0	8.2
östslaviska språk – engelska	38.7	20.8
<b>slaviska språk – engelska</b>	<b>42.7</b>	<b>22.9</b>
indoeuropeiska språk – engelska	41.7	18.1

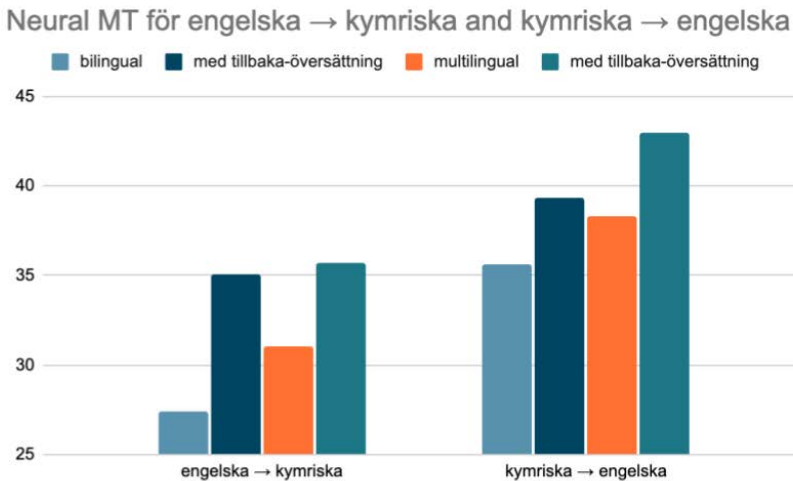
Vi kan se att flerspråkiga modeller leder till en markant förbättring jämfört med den tvåspråkiga grundmodellen (första raden). Dessutom ser vi att det finns gränser för vad sådana modeller klarar av. Att inkludera ett stort antal indoeuropeiska språk blir för mycket för dem, på så sätt att kapaciteten inom modellen inte längre räcker till för att översätta våra testdata lika bra (sista raden).

Ett annat sätt att tackla resursbristen är att skapa artificiella träningsdata. Det låter kanske suspekt, men grundidén är att utnyttja inkompleta resurser för att förbättra vissa aspekter av ett översättningssystem. Det är till exempel möjligt att göra ett system mer flytande i målspråket genom att inkludera enspråkiga målspråkstexter. Dessa kan översättas automatiskt med hjälp av ett översättningssystem i den motsatta riktningen. Följande bild illustrerar denna “tillbaka-översättning” och förklarar hur idén kan leda till en iterativ metod som stegvis kan förbättra våra system.

Frågan är om det verkligen fungerar: Vi har testat denna metod med olika lågresursspråk för att se vilka resultat man kan uppnå. Följande diagram illustrerar översättningskvaliteten (i mån av BLEU värden på ett oberoende testset) när vi lägger till så kallade “back-translations” eller automatiska översättningar av enspråkiga målspråkstexter. Denna gång tar vi engelska-kymriska som exempel.



Figur 4. Att skapa träningsmaterial med iterativ "tillbaka-översättning"

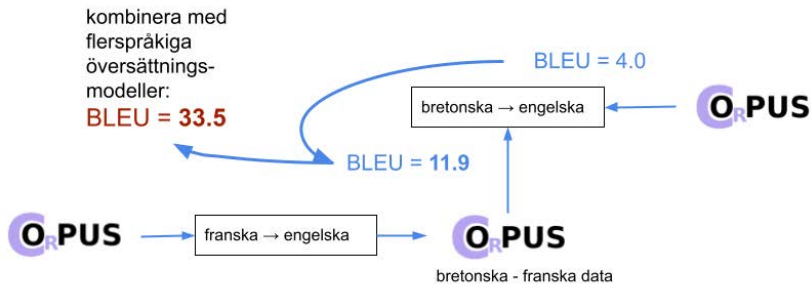


Figur 5. Effekt av automatisk datagenerering och kunskapsöverföring med relaterade språk

Den andra stapeln visar den substantiella ökning vi kan uppnå med sådana automatiskt genererade träningsdata. I synnerhet för översättningen till kymriska kan vi se en rejäl förbättring som visar vikten av språkkunskaper i målspråket särskilt för morfologiskt rika språk. Dessutom kan man lätt kombinera denna datagenereringsmetod med flerspråkiga modeller. I stapeldiagrammet kan vi se effekten av en kombination av båda metoderna med resultat för kymriska till och från engelska när vi använder modeller som är tränade på keltiska språk i stället för enbart kymriska.

Till sist vill jag också presentera ännu en metod för att utöka träningsdata. Vi tar exemplet bretonska till engelska, där en basmodell ger väldigt dåliga

resultat. Om vi mäter i BLEU igen får vi ett värde på bara 4,0 på vårt testset. Det betyder att översättningen i princip är obrukbar. Även bretonska – franska träningsdata ingår emellertid i vår databas och vi har en hyfsat bra maskinöversättning mellan franska och engelska. Det betyder att vi kan översätta franska till engelska och skapa nya automatiskt genererade exempel för vårt bretonska – engelska maskinöversättningssystem. När vi agerar så höjs kvaliteten med över 7 BLEU poäng:



Figur 6. Datagenerering med hjälp av resursrika mellanspråk

Dessutom kan vi kombinera allt detta med flerspråkiga modeller och tillämpa tillbaka-översättningsmetoden, och når till slut till ett värde på över 33 BLEU-poäng, vilket betyder att vår modell helt klart har blivit användbar.

Dessa exempel ger ett liten överblick av olika enkla metoder att hantera en situation med begränsade datamängder. Vår forskning går ut på att vidareutveckla dessa idéer och optimera inlärningsprocessen för många språk, så att vi kan nå en bättre täckning även för minoritetsspråk.

#### 4. Varför ska vi satsa på öppen maskinöversättning?

Automatisk översättning väcker en hel del diskussion. Här är några viktiga punkter att ta hänsyn till:

*“Maskinöversättning tar människors jobb”*

Ett vanligt argument mot maskinöversättning är att det förstör jobbet för mänskliga översättare och att den samtidigt leder till sämre service för små språk. Det är sant att maskinöversättning eftersträvar automatisering och övertar uppgifter som annars måste utföras av människor. Men hotet mot översättare kommer trots allt från ett annat håll. Vårt samhälle är fullt av ojorda översättningsjobb. Det finns helt enkelt ingen möjlighet att hantera den gigantiska mängden av information som borde bli tillgänglig för alla på lika villkor. Teknologiska lös-

ningar är tyvärr det enda sättet att förbättra situationen. Det kommer ändå inte finnas någon brist på uppgifter för professionella översättare där tekniken inte räcker till. Där det finns nedskärningar handlar de om politiska beslut och inte om brist på jobb.

Som en jämförelse vill jag ta upp förändringarna inom transportindustrin när bilarna kom med i bilden. Säkert var många inom den branschen rädda när de första bilarna dök upp och det kändes som om transportpersonal inte längre behövdes. Tvärtom har industrin inom logistiken bara exploderat, och jag ser likheter med översättningsbranschen. Tillsammans med teknologin kan vi nu tackla översättningsjobb som vi inte kunde drömma om tidigare för att det inte fanns resurser att ens börja med dessa. Översättningsverktyg på flera språk kommer bara att öka antalet uppdrag. Jobbet kommer såklart se annorlunda ut, på samma sätt som transportindustrin ställde om sina behov. Det här händer redan och det kommer att fortsätta i språkservicesektorn.

Den andra delen av argumentet handlar om kvaliteten som försämras när automatiska verktyg introduceras istället för enbart mänskliga översättare. Även det tycker jag beror på politiska beslut och inte på sådant teknologin förorsakar. Teknologin kan ge stöd till översättningar av hög kvalitet, men man måste naturligtvis satsa på kontroll och bearbetning. Man borde också begrunda om det inte är ännu sämre service att så mycket inte alls blir översatt när resurserna för mänskliga översättare inte räcker till. Automatöversättning kan i så fall fylla en viktig funktion genom att snabbt ge tillgång till information som annars är bortom räckhåll för en viss publik.

### *“Maskinöversättning kostar ingenting”*

Ett argument mot finansiering av offentlig utveckling av maskinöversättning är att detta görs bättre, billigare och snabbare av kommersiella aktörer och att staten och offentliga forskningsinstitut aldrig kommer att kunna konkurrera med stora globala företag och inte heller borde göra det.

De stora it-giganterna har enorma marketingstrategier där de lanserar tjänster som erbjuds smidigt på nätet utan betalning. Inte bara privatpersoner utan också myndigheter bländas av detta erbjudande och inser oftast inte hur stor investeringen är för dessa företag. Massiva resurser satsas på språkteknologi, och maskinöversättning är en av de stora tillämpningar som utvecklas. Det är därför naturligt att myndigheterna inte förstår att man behöver stora resurser för att utveckla liknande tjänster när allt ju finns “gratis” på nätet.

Det finns en bra anledning till varför internetföretag erbjuder tjänster som

verkar vara gratis, men som utelämnar användarna av dem till exploatering av företagen. Anledningen är att de vill hålla kvar sina kunder på plattformen och knyta allt till tjänster där de kan kontrollera innehållet och bombardera användarna med personlig reklam som maximerar vinsten. Dessutom kan de behändigt samla in personliga data och finslipa profilen och sina egna system. Att utelämna samhället och dess informationsbehov till vinstdrivna annonseringsföretag är väldigt riskabelt, och jag tror inte att våra beslutsfattande organ har insett vilken fara detta innebär.

Det enda sättet att undvika riskerna det här medför är enligt mig att vi så snabbt som möjligt utvecklar transparenta och säkra system som är oberoende av vinstintressen för att betjäna alla medborgare med digitala tjänster som erbjuder de funktioner som är nödvändiga för rättvis, jämlik och likvärdig behandling. Det i sin tur kräver ordentliga investeringar och inte bara skrattretande småpengar som inte bygger upp en kraftfull motståndare till de globala företag som odlar sina monopol.

För ett tag sedan utlyste statsrådets kansli i Helsingfors en anskaffning av en pålitlig automatisk översättningstjänst för fyra år framöver som skulle vara anpassad för de interna behoven och kunna hantera alla data på ett säkert sätt. Kostnadstaket fastställdes till 60 000 euro. För mig låter det som ett hån mot dem som utvecklar översättningstjänster, och det visar hur pass lite en myndighet är villig att investera i en otroligt viktig och känslig tjänst.

Vi måste äntligen erkänna att språkteknologiska lösningar är oerhört viktiga i vårt digitala informationssamhälle och att vi inte kan lägga ut känsliga områden på entreprenad till externa företag. Ordentliga resurser borde satsas på en transparent utveckling av säkra tjänster, inte minst för automatisk översättning. Det borde vara självklart och ett krav för en rättvis behandling utan språklig diskriminering.

Ett argument jag också ganska ofta får höra är att ingen ändå kan konkurrera med de stora it-jättarna och att det är bortkastade pengar att investera i egna produkter. Det är synd att höra denna uppgivenhet. Inom vår forskningsgrupp har vi satsat en hel del på att visa vad vi kan uppnå även utan ordentligt statligt stöd. Öppna lösningar kan fungera. Som avslutning vill jag därför presentera vårt projekt OPUS-MT, maskinöversättning som är tillgänglig för alla och som enkelt kan integreras i professionella arbetsflöden.

## **5. Öppen maskinöversättning med OPUS-MT**

Under 15 år har vi samlat in översatta texter för att skapa världens största databas av öppna parallella data med länkade översättningstexter som i nuläget omfattar hundratals språk med totalt över tio miljarder meningar. Denna gigan-

tiska textkorpus finns tillgänglig via OPUS<sup>2</sup> och används flitigt inom språkteknologi och översättningsvetenskap.

För ungefär två år sedan satte vi igång med ett projekt som bygger på våra data och fokuserar på utvecklingen av öppna översättningsmodeller och verktyg som kan användas som en allmän resurs inom forskning och utveckling. Det började med ett projekt som fokuserade på finska och svenska inom ramen för ett pilotprojekt som generöst finansierades av Svenska kulturfonden i Finland under namnet Fiskmö<sup>3</sup> (Tiedemann et. al, 2020). Den väldigt framgångsrika utvecklingen motiverade oss att utvidga omfånget och satsa på många språk och en bred täckning av olika översättningstjänster. Det blev början till OPUS-MT<sup>4</sup> (Tiedemann & Thottingal, 2020) som redan nu har skapat över 1 000 öppet tillgängliga översättningsmodeller samt verktyg som gör det enkelt att integrera dem i praktiska tillämpningar och professionella arbetsflöden. Efter Fiskmö fick vi finansiering från den europeiska språkteknologiinfrastrukturen ELG som ett av deras pilotprojekt.<sup>5</sup> Det gjorde det möjligt att finslipa våra verktyg och förbättra integrationen av OPUS-MT inom populära översättningsplattformar. OPUS-CAT<sup>6</sup> (Nieminen, 2021) har blivit ett stabilt paket som ger professionella översättare tillgång till alla våra system enkelt via mjukvarupaket som kan installeras på vanliga personaldatorer.

Det som är speciellt med OPUS-CAT är att översättningen sker lokalt på användarens dator och att man därmed undviker alla säkerhetsproblem som uppstår med andra tjänster. Inga data lämnar den lokala omgivningen för automatisk översättning. Användaren behöver inte heller vara uppkopplad till en onlinetjänst för att använda våra verktyg. Det är en transparent lösning som utgör en motsats till de vinstdrivna program som försöker exploatera användarens information eller binda kunderna till onlinetjänster där villkor och kostnader snabbt kan förändras. Vi är övertygade om att OPUS-MT har kommit en bra bit på vägen att frigöra användarna från typiska marknadsstrategier där datasäkerhet och integritet är tvivelaktiga och oroväckande. Vi ser OPUS-MT som ett bra exempel på att det går att utveckla översättningsteknologi för allmännyttan och att nationella initiativ kan lyckas.

---

2 <https://opus.nlpl.eu/>

3 <https://blogs.helsinki.fi/fiskmo-project/>

4 <https://github.com/Helsinki-NLP/OPUS-MT>

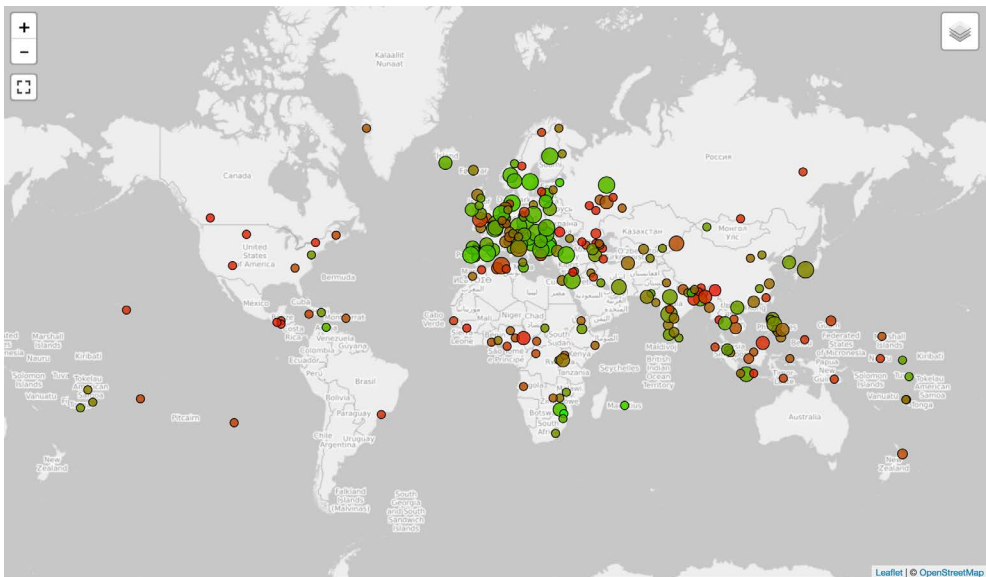
5 <https://live.european-language-grid.eu/catalogue/project/2866>

6 <https://helsinki-nlp.github.io/OPUS-CAT/>



OPUS-MT är nu tillgängligt via öppna programbibliotek<sup>7</sup> och gratis nedladdningstjänster och all källkod är öppen och transparent. Det har blivit en del av den gemensamma ansträngningen att demokratisera språkteknologiska lösningar så att vi minskar vårt beroende av stora it-företag som monopoliserar digitala tjänster. Vi hoppas på en större satsning på språkteknologisk forskning och akademisk utveckling, så att vi kan fortsätta på den här vägen som är nödvändig för en rättvis hantering av den internationella och mångspråkiga digitala världen.

Motiveringen liknar den bakom världens största encyklopedi Wikipedia: Gemensamt kan vi skapa kollektiv nytta och tackla den stora uppgiften att göra kunskap och information tillgänglig för alla. Det är fortfarande mycket som fattas innan vi är där och vi vet att bara en bråkdel av världens språk kan dra nytta av de stora framsteg som görs inom samhället. Med OPUS-MT satsar vi på mångspråkigt stöd, även om vi fortfarande är mycket begränsade.



Figur 7. Täckning av befintliga översättningsmodeller från OPUS-MT för översättning till engelska

Kartan ovan är en geografisk illustration av täckningen med översättningsmodeller från OPUS-MT för att översätta till engelska. Större cirklar betyder bättre

<sup>7</sup> Se t.ex. integrationen i transformer programbiblioteket av huggingface på <https://huggingface.co/Helsinki-NLP>

resurser. Den gröna färgen symboliserar bättre översättningskvalitet medan rött betyder en otillräcklig kvalitet.

Vi samarbetar förövrigt gärna kring förbättrade översättningstjänster för ressursfattiga språk (se t.ex. Vázquez et. al, 2021; Aulamo et. al, 2021). Vi tar också emot donationer av text eller hjälp med inläring av våra system. Ta gärna kontakt med oss på Helsingfors universitet.

## 6. Slutsatser

Språkteknologi har kommit en lång väg och praktiska tillämpningar som bygger på språkteknologi finns med när vi surfar på nätet, kollar våra informationskanaler eller använder digitala tjänster. Automatisk översättning är ett måste för att kunna hantera informationstillgång över språkliga gränser och utvecklingen visar vilken nytta vi kan få genom en integration av maskinöversättning i många olika sammanhang. Likvärdig tillgång till information är en förutsättning i en rättvis värld och det krävs stora satsningar för att ge alla världens språk tillräckligt med stöd.

Att överlåta språkhantering till stora vinstdrivna it-företag är mycket riskabelt och vi borde inse att det behövs en strategi för att bygga ett digitalt informationssamhälle där vi inte överlämnar våra privata data villkorslöst till en hungrig annonsmarknad. Därför behöver vi öppna och transparenta lösningar.

## Summary

Our society is full of unfinished translation jobs. Without technological solutions, there is simply no way to handle the gigantic amount of information that need to be made available to everyone. The article presents initiatives based on open and secure translation tools that take into account the privacy of users. Our society must support such activities, in order to enable equality without exploitation and we need to develop transparent services for barrier-free information access especially for less commercially interesting minorities and users with certain disabilities.

## Författaren

**Jörg Tiedemann** är professor i språkteknologi vid institutionen för digitala humaniora vid Helsingfors universitet. Han doktorerade inom datorlingvistik på Uppsala universitet med forskning om automatisk länkning av översatta texter och maskinöversättning innan han flyttade till universitetet i Groningen i Nederländerna för ett postdokprojekt om informationsextraktion och frågebesvarande system. Hans forskningsintressen berör huvudsakligen datadrivna metoder inom språkteknologin och han leder för närvarande ett EU-finanserat projekt om semantisk representationsinlärning och naturlig språkförståelse.



Website: <https://blogs.helsinki.fi/tiedeman/>

## Referenser

- Aulamo, M., Virpioja, S., Scherrer, Y. & Tiedemann, J., May 2021: Boosting Neural Machine Translation from Finnish to Northern Sámi with Rule-Based Backtranslation, Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa). Dobnik, S. & Øvrelid, L. (eds.). Linköping: Linköping University Electronic Press, p. 351-356 (Linköping Electronic Conference Proceedings ; no. 78)(NEALT Proceedings Series ; no. 45).
- Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z., Thorat, N., Viégas, F., Wattenberg, M., Corrado, G., Hughes, M., Dean, J., 2017: Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. In Transactions of the Association for Computational Linguistics (TACL), Volume 5, p. 339–351
- Nieminen, T.: OPUS-CAT: Desktop NMT with CAT integration and local fine-tuning. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics, 2021: System Demonstrations, p. 288–294
- Papineni, K., Roukos, S., Ward, T., Wei-Jing Zhu, 2002: Bleu: a Method for Automatic Evaluation of Machine Translation, In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, p. 311–318

- Tiedemann, J. & Thottingal, S., 1 Nov 2020: OPUS-MT -- Building open translation services for the World,, Proceedings of the 22nd Annual Conference of the European Association for Machine Translation. Martins [et al.], A. (ed.). Geneva: European Association for Machine Translation, p. 479-480
- Tiedemann, J., Nieminen, T., Aulamo, M., Kanerva, J., Leino, A., Ginter, F. & Papula, N., 1 May 2020: The FISKMÖ Project: Resources and Tools for Finnish-Swedish Machine Translation and Cross-Linguistic Research, Proceedings of the 12th Language Resources and Evaluation Conference. Calzolari, N., Béchet, F., Blache, P., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J. & Piperidis, S. (eds.). Paris: European Language Resources Association (ELRA), p. 3808-3815
- Tiedemann, J., 1 Nov 2020: The Tatoeba Translation Challenge - Realistic Data Sets for Low Resource and Multilingual MT, Proceedings of the Fifth Conference on Machine Translation. Barrault [et al.], L. (ed.). Stroudsburg: The Association for Computational Linguistics, p. 1174-1182
- Tiedemann, J., 2016: OPUS -- Parallel Corpora for Everyone, In: Baltic Journal of Modern Computing. p. 384
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I., 2017: Attention is all you need. CoRR abs/1706.03762. <http://arxiv.org/abs/1706.03762>.
- Vázquez, R., Scherrer, Y., Virpioja, S. & Tiedemann, J., 1 Jun 2021: The Helsinki submission to the AmericasNLP shared task, Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas. Mager [et al.], M. (ed.). Stroudsburg: The Association for Computational Linguistics, p. 255-264

## Nyckelord

maskinöversättning, språkteknologi, informationstillgänglighet

# Automatisk målgruppsanpassad textanpassning

*Evelina Rennes & Arne Jönsson  
Institutionen för datavetenskap  
Linköpings universitet, Linköping, Sverige*

*Språkteknologisk forskning gör det möjligt att utveckla teknik som automatiskt kan anpassa texter på olika sätt, exempelvis genom att sammanfatta eller att skriva om texter så att de blir enklare. Det är också möjligt att mäta texters komplexitet på olika nivåer genom att exempelvis mäta andelen svåra ord, meningskomplexitet och komplexitet för hela texten. Eftersom det inte finns en anpassad text som passar alla är det viktigt att göra teknikerna tillgängliga för såväl skribenter som läsare, så att de kan anpassa texten efter sin målgrupps behov. I artikeln presenterar vi ett antal tekniker för och utmaningar med detta. Vi illustrerar med exempel från system för mätning av textkomplexitet och olika typer av textanpassning.*

## 1. Inledning

Att kunna läsa och förstå skriven text är i mångt och mycket en förutsättning för ett liv där man till fullo inkluderas i samhället. Detta gäller inte minst de senaste decennierna när den digitala utvecklingen har tagit ett jättesprång, och alltmer text produceras och görs tillgänglig på internet. Faktum är att läsning ses som en så pass viktig förmåga att det nämns i FN:s deklaration om mänskliga rättigheter. I *Konvention om rättigheter för personer med funktionsnedsättning*, Artikel 21, *Yttrandefrihet och åsiktsfrihet samt tillgång till information*, står det:

Konventionsstaterna ska vidta alla ändamålsenliga åtgärder för att säkerställa att personer med funktionsnedsättning kan utöva yttrandefriheten och åsiktsfriheten, inklusive friheten att söka, ta emot och sprida uppgifter och idéer på lika villkor som andra och genom alla former av kommunikation som de själva väljer enligt definitionen i artikel 2, däribland genom att

- a) utan dröjsmål och extra kostnader förse personer med funktionsnedsättning med information som är avsedd för allmänheten i tillgängligt format och teknologi anpassad för olika funktionshinder, (SÖ 2008:26, s. 18)

Personer med olika typer av funktionsnedsättningar ska alltså ha möjlighet att kommunicera och ta till sig information på samma villkor som en person som inte har någon typ av läsproblem. Kort sagt: läsning är en mänsklig rättighet.

Även på nationell nivå är det sedan 2009 lagstadgat i språklagen (SFS 2009:600, § 11) att *"Språket i offentlig verksamhet ska vara vårdat, enkelt och begripligt"* och detta ställer naturligtvis stora krav på att myndighetstexter ska vara så begripliga som möjligt för att informationen ska nå så många som möjligt.

## 2. Lättläst och klarspråk

Lättläst och klarspråk är två begrepp som ofta, lite slarvigt, blandas ihop. På engelska är det än värre, där forskare och skribenter använder olika uttryck som Easy Language, Easy-to-Read, Easy Reading, Easy to Understand, Simplified Language, Clear Writing och Plain Language.

*Klarspråk*, det som på engelska ofta benämns som Plain Language, handlar om att myndighetstexter ska vara skrivna på ett *vårdat, enkelt och begripligt* språk (SFS 2009:600, § 11) för att uppnå målet att inkludera så många läsare som möjligt. Begreppet *lättläst* (MTM, 2021) innebär ett tydligare målgruppsfokus och inkluderar alla typer av texter. Lättlästa nyheter och lättlästa skönlitterära böcker är exempel på texter som faller under lättlästskategorin. I den här artikeln använder vi oss av båda begreppen, men även *lätt svenska* som här fungerar som ett paraplybegrepp för de båda termerna.

För att säkerställa att text skrivs på ett så enkelt och begripligt sätt som möjligt har det tagits fram riktlinjer för att skriva lättläst. I Sverige har Myndigheten för Tillgängliga Medier (MTM) tagit fram riktlinjer som innefattar både rent språkliga råd om ordval och grammatik, men också råd om bildsättning, textstruktur, radavstånd och typsnitt (Myndigheten för Tillgängliga Medier (MTM), 2021). Språkrådet har även gett ut en mer omfattande guide till klarspråk för myndigheter (Språkrådet, 2014).

Eftersom riktlinjerna för att skriva begriplig text försöker rikta sig till så många som möjligt kan de av naturliga skäl upplevas ganska generella och svepande. Denna breda ansats har givetvis fördelar, bland annat blir det enklare att träna upp nya skribenter om det bara finns en begränsad mängd enhetliga riktlinjer att förhålla sig till, men det finns även en del nackdelar. Den tyngsta är att det inte finns en text som passar alla läsare. Målgruppen för begripliga texter innefattar allt från personer med dyslexi, afasi, autism, kognitiv funktionsnedsättning till personer med annat modersmål än svenska, och dessa skilda målgrupper har väldigt olika behov och upplevda svårigheter. För att komplicera det ytterligare är det inte heller säkert att individer inom en målgrupp har samma behov.

För att kompensera för denna breda ansats tydliggörs det i olika riktlinjer att det är viktigt att ta reda på vem mottagaren är, och vilka behov mottagaren har, men i verkligheten kan texterna som skrivs ha flera möjliga mottagare. Finns det då tid och resurser för att anpassa texten till alla dessa mottagare?

Språkrådet utvärderar regelbundet klarspråksanvändningen hos svenska myndigheter, kommuner och regioner med hjälp av enkäter och intervjuer av klarspråksansvariga skribenter. I 2020 års version av denna utvärdering kom man bland annat fram till att det främsta hindret för att tillhandahålla information på klarspråk är tidsbrist (Hansson, 2020).

En möjlig lösning på problemet med tidsbrist är en effektiv teknisk lösning för att producera enkel text, en teknisk lösning som dessutom kan anpassa texterna efter personer med olika typer av läsbeteende och kanske till och med erbjuda en individuell anpassning.

### 3. Automatisk textanpassning för olika målgrupper

Att utveckla teknik för att automatiskt anpassa texter till målgrupper eller enskilda personer innebär en mängd utmaningar. Den största utmaningen handlar om att få grepp om läsaren. Har hen problem med figurativt språk, behöver hen få förklaringar på kulturella fenomen eller svåra ord? Har hen begränsat arbetsminne och behöver korta meningar med enkel syntaktisk struktur?

Vi vet en del om olika grupperns läsproblem. I tabell 1 ges en kortfattad sammanfattning av några typiska upplevda svårigheter hos olika grupper, men skillnaderna mellan olika individer är stora och den enskilde individen kan dessutom tillhöra fler än en målgrupp.

Tabell 1. Exempel på olika målgruppers svårigheter

Målgrupp	Exempel på upplevda svårigheter
Dyslexi	Långa och ovanliga ord, homofoner, ord som är ortografiskt lika, nya ord
Afasi	Hög informationsdensitet, långa meningar, långa sekvenser av adjektiv, passiv form, sammansatta ord
Andraspråksinläring	Förståelse av ord och kulturella företeelser, textstruktur
Hörselskada	Komplexa grammatiska konstruktioner, textstruktur, långa meningar, förståelse av ord
Intellektuell funktionsnedsättning	Svårigheter relaterade till arbetsminnet, motivation till läsning, avkodning av text, textförståelse

Ytterligare utmaningar är att veta i vilken grad texter ska anpassas till respektive läsargrupp, i vilka sammanhang som anpassningar ska ske och med vilket

format de ska presenteras för läsaren. Dessutom är det viktigt att förenklingarna inte innebär att relevant information tas bort eller ändras. Det finns alltså en mängd delproblem som inte är triviala.

De språkteknologiska teknikerna för textanpassning utvecklas dock i rask takt, och verktyg som på olika sätt tar sig an problemet att öka den digitala inkluderingen görs i högre grad tillgängliga. Grovt kan de delas in i tre kategorier:

- **Tekniker för att mäta en texts komplexitet.** Dessa kan användas för att veta vilka egenskaper en text har och för att hitta texter som passar en viss grupp av läsare, eller till att förenkla texter med en viss grad av komplexitet.
- **Tekniker för att sammanfatta en text.** Här skiljer man vanligtvis mellan extraktiv och abstraktiv sammanfattning. I den extraktiva sammanfattningen plockas de viktigaste meningarna ut, i den abstraktiva sammanfattningen skrivs texten om helt.
- **Tekniker för att förenkla en text.** Detta kan bland annat ske syntaktiskt, genom att skriva om grammatiskt svåra meningar till enklare, eller lexikalt, där svåra ord och fraser byts ut mot enklare synonymer.

#### 4. Textkomplexitetsmätning

Om syftet är att anpassa en text så att den blir enklare behöver man först ta reda på hur svår texten är från början, och om det finns vissa komponenter i texten som är onödigt komplexa. Det kan göras med hjälp av en mer eller mindre sofistikerad komplexitetsmätning.

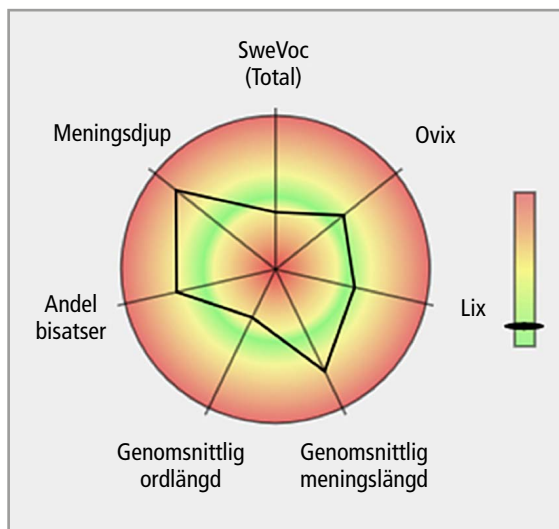
Dagens språkteknologiska tekniker låter oss nämligen att mäta det mesta i en text, och de mått som finns kan grovt delas upp i följande kategorier (Falkenjack et al., 2017):

- Ytliga mått
  - Mått som är enkla att räkna ut och inte behöver någon djupare analys, exempelvis antal ord per mening. Långa meningar är ofta svårare att förstå.
- Lexikala mått
  - Mått som är baserade på ordfrekvenser och grundläggande svensk vokabulär. Här ger ordlistorna från SweVoc (Heimann Mühlenbock & Johansson Kokkinakis, 2012) en bra grund för att ange vilka ord som är vanliga och vilka som är mindre vanliga. Ett vanligt och vardagligt ord upplevs oftare som enklare att läsa.



- Morfosyntaktiska mått
  - Mått som bygger på en morfologisk analys av texten och som gör att vi kan räkna exempelvis antalet innehållsord såsom substantiv, verb, adjektiv och adverb i en text. En hög andel indikerar en innehållsrik text som kan vara svår för individer med vissa typer av lässvårigheter. Å andra sidan kan en låg andel innehållsord indikera en syntaktiskt komplex text, vilket i sig kan utgöra en svårighet.
- Syntaktiska mått
  - Mått som speglar egenskaper beräknade efter en syntaktisk analys av texten. Hit hör bland annat dependenslängden, det vill säga avståndet mellan ett ord och dess huvud, oftast meningens verb. Texter med stora avstånd mellan sammanhängande ord är svårare att förstå än texter med mindre avstånd.
- Traditionella textkvalitetsmått
  - Mått som traditionellt har använts för att mäta läsbarhet, exempelvis LIX och OVIX.

Textkomplexitetsmått ger alltså en möjlighet att mäta en texts komplexitet utifrån flera perspektiv. För att måtten ska vara verkligt användbara krävs att de görs begripliga och går att koppla till individers upplevda läsning. Här ställs man inför två problem. Det första problemet är hur man ska tolka resultaten, som ofta kommer tillbaka som en tabell av siffror. En möjlighet skulle kunna vara att visualisera komplexiteten, exempelvis som i figur 1, i analogi med



Figur 1. Exempel på visualisering av textkomplexitet

Santini (Santini et al., 2020). Här behövs betydligt mer forskning kring vilka mått som bör visualiseras och även *hur* de bör visualiseras för att hjälpa en användare snarare än att stjälpa.

Det andra problemet är att koppla måtten till individers eller grupperas faktiska läsförmåga. Det vi ser i textkomplexitetsmått är själva textens egenskaper, som egentligen inte säger så mycket om hur en läsare kommer att uppleva läsningen av texten.

## 5. Automatisk textsammanfattning

Att erbjuda en läsare en sammanfattning kan underlätta genom att den viktigaste informationen i en text tydliggörs och samtidigt kan texten kännas kortare och enklare att ta till sig. Det finns två olika tekniker för att göra detta, extraktiv och abstraktiv sammanfattning. Extraktiv sammanfattning fungerar genom att de viktigaste meningarna plockas ut direkt från originaltexten och bildar en ny text, medan den abstraktiva sammanfattningen genererar en helt ny text som är en sammanfattning av originaltexten.

Antag att vi har följande, ganska långa, originaltext:

Tillsammans med Kanadas nye stjärna Sidney Crosby duellerade 20-åriga Alexander Ovetjkin om vem som skulle bli årets rookie i NHL. Poängmässigt vann den tuffe och spelskicklige ryssen i Capitals, laget från USA:s huvudstad. På 81 matcher under sin första säsong i NHL gjorde han 52 mål och passade till 54. 106 poäng en rookiesäsong är bland det absolut bästa en nykomling presterat i ligan. Han debuterade redan som 16-åring i den ryska ligan och trots locktoner om feta dollarbuntar från andra sidan av Atlanten stannade han kvar hemma i Moskva fram till och med förra säsongen. Det gjorde också att Ovetjkin var en redan klar elitspelare när han landade i Washington, där han för övrigt gjorde mål redan i sin debutmatch. Washington som lag räckte dock inte alls till för att ta sig till årets Stanley Cup-slutspel. Det är därför Alexander Ovetjkin nu kan komma och visa upp sig i Globen redan i kväll och därefter förstärka det ryska VM-laget i Riga. Det är alla fall var den ryska lagledningen hoppas på. Ovetjkin skulle egentligen ha anslutit till det ryska laget som på torsdagseftermiddagen kom från Helsingfors. Ryssland spelade sin första match i Hockey Games mot Finland i Helsingfors och vann i onsdags kväll. Nu blev Ovetjkin försenad från USA och landar inte i Stockholm förrän på fredagsmorgonen. Alexander Ovetjkin är bara en i raden av unga framgångsrika ryska hockeyspelare som kommer fram just nu. Tillsammans med bland andra Ilja Kovaltjuk, Atlanta och Jevgenij

Malkin, Magnitogorsk i ryska ligan, har rysk hockey fått fram stjärnor som kan ta tillbaka landets hockeylandslag till en nivå vi inte sett sedan början av 1990-talet. Med Ovetjkin i laget blir det ännu svårare för Sverige att dels vinna dagens match i Globen och även att ta hem slutsegern i Euro Hockey Tour, som avgörs på måndag.

En extraktiv sammanfattning där man alltså plockar ut de viktigaste meningarna skulle då kunna se ut så här:

Det är därför Alexander Ovetjkin nu kan komma och visa upp sig i Globen redan i kväll och därefter förstärka det ryska VM-laget i Riga. Alexander Ovetjkin är bara en i raden av unga framgångsrika ryska hockeyspelare som kommer fram just nu. Tillsammans med Kanadas nye stjärna Sidney Crosby duellerade 20-åriga Alexander Ovetjkin om vem som skulle bli årets rookie i NHL. Med Ovetjkin i laget blir det ännu svårare för Sverige att dels vinna dagens match i Globen och även att ta hem slutsegern i Euro Hockey Tour, som avgörs på måndag.

Den extraktiva sammanfattaren har producerat en betydligt kortare text genom att plocka ut de viktigaste meningarna. Tekniken att göra detta bygger på att man skapar ordvektorer och representerar texten i ett flerdimensionellt vektorrum. Ordvektorer kan skapas på flera olika sätt men grundtanken är att ett ords betydelse beror av det sammanhang i vilket det används. Genom att till exempel tilldela varje ord en slumpmässig vektor av hög dimension, några hundra är vanligt, och sedan till varje ords vektor addera dess närmsta grannars ordvektorer kommer ordet att få en ny vektor som beror av det sammanhang i vilket ordet användes. Genom att upprepa detta tusentals gånger för varje ord i texter om miljontals ord kommer så småningom ordvektorerna att inte längre vara slumpmässiga. Ordvektormodeller finns fritt tillgängliga, för svenska t.ex. KB-BERT (Malmsten et al., 2020).

För att skapa en sammanfattning skapar man en vektorrepresentation för hela texten och varje mening för sig och sen väljer de meningar vars vektorer ligger närmast originaltexten. Eftersom orden är representerade som vektorer kan man använda cosinus för att direkt mäta hur nära meningarna ligger texten. Man tar också hänsyn till vilka koncept i texten som är viktigast och vilka meningar som innehåller detta, se exempelvis (Smith & Jönsson, 2011).

Den abstraktiva sammanfattaren, som tränats på en stor mängd redan sammanfattade texter, skriver i stället en sammanfattning med nya ord. Att hitta en stor mängd av korrekt sammanfattade texter som kan användas till träning

är svårt, men om man antar att ingressen till en nyhetsartikel sammanfattar brödtexten kan dessa användas som träningsdata (Monsen & Jönsson, 2021). I den abstraktiva sammanfattaren används återigen ordvektorer. Mer specifikt har modellen tränats för att känna igen vilken information i en given text som är relevant att ha med i sammanfattningen och, givet denna information, predicera vilket ord som följer på ett annat. När en ny text sedan matas in i modellen kommer den mest sannolika sammanfattningen, baserat på den tränade modellen, att matas ut. Ett exempel på detta, är följande:

Alexander Ovetjkin är klar för spel i det ryska VM-laget i ishockey. Det innebär att han kommer att förstärka det ryska laget i kvällens match mot Sverige i Globen. Det blir Ovetjkins första match i Moskva och den blir tuffare än vad han hade tänkt sig innan han lämnade Washington Capitals för att spela i ryska ligan. Det är nu ryska hockeyligan, NHL, är på väg att ta upp sporten på den nivå man inte sett sedan början av 1990-talet. I dag möts Sverige och Ryssland i Globen i en match som kan bli avgörande för om Sverige ska vinna Euro Hockey Tour.

Här ser vi att meningarna är helt omskrivna, och modellen har skapat en helt ny text. Vi ser också att meningarna är grammatiskt korrekta men att innehållet skiljer sig något från det som genererats av den extraktiva sammanfattaren. Vilken sammanfattning som objektivt sett är bäst är inte lätt att säga.

Fördelarna med den extraktiva sammanfattningen är att den är relativt enkel, rent metodmässigt, och att sammanfattningen inte adderar någon ny information. Nackdelarna med den extraktiva sammanfattningen är att texten kan upplevas som lite stolpig och att relevant information kan gå förlorad i extraktionsprocessen. I exemplet har detta skett när första meningen inleds med ett "därför" utan att det finns någon mening tidigare som motiverar detta "därför".

Den abstraktiva sammanfattningen har å andra sidan fördelen att det blir ett mer naturligt flyt genom texten, vilket kan förhöja läsbarheten och läsoplevelsen, men nackdelen att felaktig information riskerar att läggas till texten. I exemplet syns detta t.ex. i det felaktiga "Det är nu ryska hockeyligan, NHL".

Vi har i studier visat att den extraktiva sammanfattningstekniken i viss mån föredras om man presenteras för texter sammanfattade med hjälp av båda teknikerna samt originaltexten.

## **6. Automatisk textförenkling**

Att göra en text syntaktiskt, lexikalt eller semantiskt enklare med hjälp av språkteknologiska verktyg kallas för automatisk textförenkling. Det kan bland

annat göras genom att byta ut ord i texten mot enklare synonymer, lägga till en förklaring till ett svårt begrepp, skriva om en mening så att den får en rakare ordföljd, eller ta bort onödiga inskjutna bisatser.

Automatiskt byte av synonymer görs relativt enkelt. Det finns språkliga resurser att tillgå på de flesta språk som möjliggör att ord kan slås upp och bytas ut mot andra. Däremot är det fortfarande en öppen fråga hur synonymerna ska väljas ut. Här kan studier av hur professionella skribenter gör för att förenkla sin text ge värdefull kunskap.

Synonymutbyten kan göras på olika sätt, exempelvis genom att byta ut ett ord baserat på frekvens, där vanligare ord ofta är enklare att förstå. Till exempel är *lön* (frekvens 1057 i den svenska Parole-ordlistan<sup>1</sup>) enklare att förstå för de flesta än *inkomst* (frekvens 378). Detta fungerar dock inte alltid, till exempel är *bustrun* (frekvens 608) betydligt vanligare än *frun* (frekvens 213). Då kan man i stället, eller som ett komplement, välja andra strategier för att hitta den mest passade synonymen. Sådana strategier kan vara att se till ordets längd och välja det kortaste ordet som synonym, eller att utgå från andra välstuderade egenskaper hos orden så som konkretionsgrad eller vid vilken ålder som ordet vanligen lärs in.

*Synlex* (Kann, 2004) är en resurs där föreslagna synonympar annoterats med grad av synonymitet, det vill säga, där personer har fått betygsätta hur synonyma två ord är. Men även här kan det bli svårt att välja rätt synonym. I *Synlex* listas till exempel *ersättning*, *gage*, *honorar* och *lön* som möjliga synonymer till *arvode* och med snarlika grader av synonymitet. Här kan sammanhanget möjligen hjälpa till.

Slutligen visar det sig att *basord*, eller vad som ofta kallas *prototyper*, ofta är enklare att förstå (Rennes & Jönsson, 2021), till exempel är *bund* enklare att förstå än *terrier* eller *dvärgschnauzer*. Men, precis som vid de andra strategierna för synonymbyte kan en del precision gå förlorad om man väljer basordet till förmån för det mer specifika ordet.

Automatisk omskrivning av komplicerade meningar till enklare kan göras på i princip två sätt, antingen genom att skriva regler som anger hur en mening kan förenklas, eller genom att träna modeller på parallellställda korpusar, i princip på samma sätt som maskinöversättning fungerar. Det finns för- och nackdelar med de olika metoderna. Regelbaserade system tillåter mer precision och skräddarsydda omskrivningar, men det är tidskrävande att konstruera regler som fungerar i alla sammanhang. Modellbaserade system kräver inte mycket manuellt arbete, men däremot behövs en större mängd högkvalitativa

---

1 <https://spraakbanken.gu.se/resurser/parolelex>

och parallella data för att prestera ett bra slutresultat, vilket visar på vikten av data för textanpassning.

Regelbaserade textförenklningar bygger på att identifiera vissa språkliga konstruktioner i meningar och sedan skriva om dem. En regel som till exempel skriver om från passiv form till aktiv måste först göra subjektet i den passiva meningen till objekt i den aktiva. Därefter måste s-ändelsen på verbet tas bort och eventuell agent transformeras till subjekt. Meningen: *Den enorma kakan åts av både Kalle och Stina i matsalen* kommer då att skrivas om till: *Både Kalle och Stina åt den enorma kakan i matsalen*. Andra förenklingsregler kan handla om att tillämpa rak ordföljd (*Igår köpte Kalle en ny bil* → *Kalle köpte en ny bil igår*), att dela upp långa i flera kortare och citatomvandling ("*Gå och lägg dig!*" sade Kalle → *Kalle sade: "Gå och lägg dig!"*).

Det finns förstås väldigt många möjliga förenklingsregler och en del fungerar genom att ta bort information som kanske kan vara onödig, till exempel vissa adjektiv (*Kalle har en stor blågrön bil* → *Kalle har en bil*). Problemet här är att inte ta bort för mycket som följande exempel från ett textförenklingsystem visar: *Europa är en av de mest urbaniserade kontinenterna i världen* → *Europa är en av kontinenterna*.

## 7. Språkliga data

Professionella skribenter har under många år anpassat texter till olika målgrupper. Sådana texter kan vi analysera och använda oss av för att få en förståelse av hur vi bör förenkla texter, vilket i sin tur kan användas till skrivstöd för skribenter eller för att utveckla de tekniker för automatisk textanpassning som beskrivits ovan.

För att utveckla olika tekniker för automatisk textanpassning kan korpusar av förenklade texter fungera som facit, en guldstandard, dit man vill nå med sina textförenklningar. Detta är användbart för att utvärdera regelbaserad teknik och är nödvändigt för modellbaserade tekniker som bygger på att man använder maskininlärning för att träna en modell på data som anses representera det man vill uppnå. Den abstraktiva sammanfattaren som presenterades ovan tränades till exempel på en korpus bestående av 38 151 artiklar från Dagens Nyheter där ingressen användes som sammanfattning av artikeln (Monsen & Jönsson, 2021).

Tanken bakom att använda textkorpusar på detta sätt är att det som de professionella skribenterna gör är korrekt och dit man vill sträva. Dessvärre finns det väldigt få texter på svenska som skrivits för en specifik målgrupp, där både de förenklade texterna och originaltexterna finns tillgängliga. I stället är korpusar på svenska med texter på lätt svenska hämtade från ett material som inte är målgruppsanpassat, eller som försöker nå en väldigt bred målgrupp.

För svenska finns det en större korpus av svenska lättlästa texter som heter LäSBarT<sup>2</sup>. LäSBarT innehåller drygt 1 miljon ord uppdelade på fyra olika genrer och har använts bland annat för att bättre förstå textförenkling och studera textkomplexitet (Falkenjack et al., 2013).

För att bygga modeller för textförenkling som beskrivits ovan vore det användbart att ha korpusar som är parallellställda, det vill säga där en mening har matchats ihop med en förenklad motsvarighet. En svensk sådan korpus innehåller samtliga svenska offentliga förvaltningars (myndigheter, kommuner och regioner) publika webbsidor från 2017 på vanlig svenska respektive lättläst svenska<sup>3</sup> (Rennes, 2018).

Tabell 2. Beskrivning av en svensk parallellställd korpus

	Vanlig svenska	Lätt svenska
Antal dokument	136 501	1629
Antal tecken	29,2 miljoner	334 491
Parallella	15 433 unika meningspar	

Korpusen beskrivs i tabell 2 där man tydligt ser att det finns väldigt få sidor på lätt svenska. Korpusen har också parallellställts genom att utgå från de enkla meningarna och leta efter meningar med motsvarande innehåll bland de texter som är skrivna på vanlig svenska (Rennes, 2020). För att räkna ut vilka meningar som har mest lika innehåll används återigen ordvektorer. Här tar man varje ordvektor i meningsparen, räknar ut hur lika de är, cosinusavståndet, och summerar detta för hela meningen. Därefter väljer man ut de meningspar vars summa överstiger en tröskel. Tröskelvärdet provas ut genom att be informanter avgöra hur lika två meningar är, en lättläst och en på vanlig svenska. Genom att göra detta för ett urval meningspar med olika cosinusavstånd får vi fram ett värde då meningsparen inte längre anses lika. Detta resulterade i 15 433 unika meningspar. Tyvärr är detta alldeles för få meningspar för att kunna träna en modell för automatisk textförenkling. Däremot kan den användas för att utveckla nya och bättre förenklingsregler.

## 8. Implementerade verktyg

Att kunna läsa och förstå skriven text är en förutsättning för att inkluderas i

2 <https://spraakbanken.gu.se/resurser/lasbart>

3 <https://www.ida.liu.se/~arnjo82/diginclude/corpus.shtml> (Rennes, An Aligned Resource of Swedish Complex-Simple Sentence Pairs, 2018)

samhället, och varje medborgare ska ha samma rätt att kunna ta till sig information. Vårt strå till stacken är en palett av tekniker och verktyg som kan användas både för att göra texter enklare att läsa och som ett skrivstöd för att skriva begripligt.

*FriendlyReader* är ett verktyg som riktar sig mot slutanvändaren. Idén med verktyget är att den enskilde läsaren själv ska kunna anpassa texten efter hans specifika behov. *FriendlyReader* innehåller i dagsläget möjligheten att sammanfatta texten, både extraktivt och abstraktivt, få synonymförslag till svåra ord och skriva om texten med enklare grammatisk struktur. Utöver de rent språkteknologiska funktionerna är det även möjligt att anpassa textstorlek, radavstånd, typsnitt och att få texten uppläst.

*TeCST* är ett verktyg som riktar sig mot skribenter och är tänkt som ett skrivstöd för att effektivt producera enkel text. *TeCST* innehåller möjligheten att få en texts komplexitet bedömd och visualiserad, att få förslag på syntaktiska förenklingar, samt få markeringar av svåra grammatiska strukturer och svåra ord i texten.

Vår intention med verktygen är att de ska kunna göra det möjligt att skraddarsy textanpassningar till en vald målgrupp, kanske till och med en enskild person. Bara genom att se till den enskilda läsaren och de utmaningar som denne ställs inför kan vi skapa texter som är tillgängliga på riktigt.

## Summary

Language technology research makes it possible to develop technology for the automatic adaptation of text to enhance readability and comprehension. This can be done by creating automatic summaries of the text, or by rewriting the text so that it contains simpler words and less complex grammatical structures. It is also possible to measure the complexity of texts from various perspectives and text levels, including the proportion of difficult words, sentence complexity and measures on the complexity for the entire text. Since there is no *one-size-fits-all* text adaptation, it is important to make the techniques available to both writers and readers so that they can adapt the text to their needs and know how complicated it is. In the article, we present several techniques and illustrate with examples from real systems for text complexity measurement and text adaptation.



## Författare

**Evelina Rennes** doktorerar i datavetenskap vid Linköpings universitet. I sin forskning intresserar hon sig framför allt för automatisk textanpassning för personer med olika typer av lässvårigheter.



**Arne Jönsson** är professor i datavetenskap vid Linköpings universitet. Han har under flera år forskat inom artificiell intelligens, de senaste åren med fokus på språkteknologi för ökad digital inkludering.



## Litteraturförteckning

- Falkenjack, Johan, Heimann Mühlenbock, Katarina & Jönsson, Arne, 2013: Features indicating readability in Swedish text. I: Stephan Oepen, Kristin Hagen, Janne Bondi Johannessen (red.): *Proceedings of the 19th Nordic Conference of Computational Linguistics (NoDaLiDa-2013)*, Oslo, Norge. s. 27–40. Linköping: Linköping University Electronic Press.
- Falkenjack, Johan o.a., 2017: Services for text simplification and analysis. I: Jörg Tiedemann, Nina Tahmasebi (red.): *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa-2017)* Göteborg, Sverige, s. 309–313. Linköping: Linköping University Electronic Press.
- Hansson, Karin, 2020: "Det finns ett sug efter klarspråk" *En studie om bättre stöd till klarspråk i offentlig verksamhet*. Stockholm: Institutet för språk och folkminnen. Tillgänglig på Internet: <https://isof.diva-portal.org/smash/get/diva2:1440648/FULLTEXT01.pdf>

- Heimann Mühlenbock, Katarina & Johansson Kokkinakis, Sofie, 2012: SweVoc – A Swedish vocabulary resource for CALL. I: Lars Borin, Elena Volodina (red.): *Proceedings of the SLTC 2012 workshop on NLP for CALL, Lund, Sverige*, s. 28–34. Linköping: Linköping University Electronic Press.
- Kann, Viggo, 2004: *Folkets användning av Lexin – en resurs*. Tillgänglig på Internet: <http://www.csc.kth.se/~viggo/rapporter/synlex.pdf>
- Malmsten, Martin, Börjeson, Love & Haffenden Chris, 2020: Playing with Words at the National Library of Sweden – Making a Swedish BERT, *arxiv.org:2007.01658v1 [cs.CL]*.
- Mühlenbock, Katarina, 2008: Readable, Legible or Plain Words – Presentation of an easy-to-read Swedish corpus. I: Anju Saxena, Åke Viberg (red.): *Multilingualism: Proceedings of the 23rd Scandinavian Conference of Linguistics*, s. 327–329. Uppsala: Acta Universitatis Upsaliensis.
- Monsen, Julius & Jönsson, Arne, 2021: A method for building non-English corpora for abstractive text summarization. I: Monica Monachini, Maria Eskevich (red.): *Proceedings of the CLARIN annual conference*, s. 82–85.
- Myndigheten för Tillgängliga Medier (MTM), 2021: *Att skriva lättläst*. Tillgänglig på Internet: <https://www.mtm.se/var-verksamhet/lattlast/att-skriva-lattlast/>
- Rennes, Evelina, 2018: An Aligned Resource of Swedish Complex-Simple Sentence Pairs. I: *Proceedings of the Seventh Swedish Language Technology Conference (SLTC), Stockholm, Sverige*, s. 61–63.
- Rennes, Evelina, 2020: Is it simpler? An Evaluation of an Aligned Corpus of Standard-Simple Sentences. I: Núria Gala, Rodrigo Wilkens (red.): *Proceedings of the LREC Workshop on Tools and Resources to Empower People with READING Difficulties (READI-20), Marseille, France*, s. 6–13. European Language Resources Association.
- Rennes, Evelina & Jönsson, Arne, 2021: Synonym Replacement based on a Study of Basic-level Nouns in Swedish Texts of Different Complexity. I: Simon Dobnik, Lilja Øvrelid (red.): *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa), Reykjavik, Island*, s. 259–267. Linköping: Linköping University Electronic Press.

- Santini, Marina, Jönsson, Arne & Rennes, Evelina, 2020: Visualizing Facets of Text Complexity across Registers. I: Núria Gala, Rodrigo Wilkens (red.): *Proceedings of the LREC workshop Tools and Resources to Empower People with READING Difficulties (READI-20)*, s. 49–56. European Language Resources Association.
- Smith, Christian & Jönsson, Arne, 2011: Enhancing extraction based summarization with outside word space. I: Haifeng Wang, David Yarowsky (red.): *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP), Chiang Mai, Thailand*, s. 1062–1070. Asian Federation of Natural Language Processing.
- Språkrådet, 2014: *Myndigheternas skrivregler, 8 upplagan*. Stockholm: Norstedts Juridik AB/Fritzes. Tillgänglig på Internet: <https://www.isof.se/download/18.17dda5f1791cdbc2873a99/1620030264840/Mynd-skrivreg2014-1.pdf>
- SÖ 2008:26, 2008: *Konvention om rättigheter för personer med funktionsnedsättning*. Tillgänglig på internet: <https://www.regeringen.se/4ae1cb/globalassets/regeringen/dokument/socialdepartementet/funktionshinder/konvention-om-rattigheter-for-personer-med-funktionsnedsattning.pdf>

## **Nyckelord**

Automatisk textanpassning, textkomplexitet, automatisk textsammanfattning, automatisk textförenkling, lässvårigheter



# Talegenkendelse og sprogteknologi på Færøerne

*Sandra Saxov Lambauge & Iben Nyholm Debess*

*Denne artikel er skrevet på baggrund af et oplæg om talegenkendelse og sprogteknologi på Færøerne på Nordisk Sprogmode 2021. Artiklen præsenterer det igangværende færøske projekt, projekt Ravn. Projektet udvikler den første færøske talegenkender, og til dette formål udvikles også en BLARK, som indeholder sproglige ressourcer egnet til sprogteknologisk arbejde generelt. Ressourcerne er open source og indeholder bl.a. en ordbog, et SAMPA-alfabet, optagelser i wav-format og et tekstkorpus. Med afsæt i det aktuelle projekt giver vi et overblik over sprogteknologien på Færøerne generelt.*

## 1. Projekt Ravn

Da projekt Ravn begyndte i 2019, fandtes der næsten ingen sprogressourcer egnede til sprogteknologi for det færøske sprog, hverken ordbøger, taledata eller tekstkorpora. Derfor var det en nødvendighed for os at udvikle sproglige ressourcer af høj kvalitet, som kunne bruges både til vores talegenkender og til færøsk sprogteknologi generelt.

Det igangværende projekt Ravn er et uafhængigt projekt, der er økonomisk forankret i den færøske fond Talutøkni. Dette treårige projekt blev søsat i januar 2019, og afslutningen er projekteret til 2022. Projektet er finansieret af det færøske kulturministerium, nogle færøske kommuner, Folketinget og færøske private virksomheder og foreninger. Projektet er ikke tilknyttet nogen offentlig institution eller privat virksomhed, og alt det materiale, som projektet resulterer i, bliver udviklet uden tanke på profit eller ejerrettigheder.

Projektets primære formål er at udvikle talegenkendelse til færøsk. Talegenkendelse er en automatisk afkodning af naturligt sprog ved hjælp af et program - et program, der gør det muligt for en computer, telefon eller anden platform at skrive og ultimativt forstå det, der bliver sagt. Da færøsk i forvejen har talesyntese (et program, der omdanner tekst på en computer/telefon til tale), var det et naturligt næste skridt at gå i gang med talegenkendelse. Disse to taleteknologiske værktøjer er meget nyttige for forskellige brugergrupper, både ved privat og professionelt brug. Den færøske talesyntese er velimplementeret på mange platforme og bliver flittigt brugt, og mange af brugerne efterspørger at kunne bruge taleteknologien den anden vej rundt.

Ved projektets start var det faktisk nødvendigt for arbejdsgruppen at finde ud af, hvad en 'talegenkender' skulle hedde på færøsk - manglen på terminologi illustrerer, hvor lidt aktivitet der havde været med talegenkendelse tidligere. Dermed vil dette projekt udvikle den første færøske 'talukennari'.

Projekt Ravns sekundære formål er at udvikle sprogressourcer egnet til sprogteknologisk arbejde. Det har været et vilkår for dette projekt, at der har været sparsomt med færøksproglige ressourcer, der er egnede til arbejde med talegenkendelse. Selvom der findes gode færøske sprogressourcer, er nytteværdien begrænset af årsager som teknisk format, adgang og rettigheder. Derfor består projektets konkrete arbejde i at udvikle sprogligt materiale fra bunden i god kvalitet, i åbne formater og med åben adgang.

Dette sproglige arbejde går selvfølgelig hånd i hånd med den anden side af et sprogteknologisk projekt: teknologien. Sprogressourcerne skal både være sprogligt og teknisk velovervejede. Forfatterne er begge to lingvister, og derfor er fokus i denne artikel sprogligt.

## **2. BLARK**

I et lille sprogsamfund som Færøerne er der meget få sproglige ressourcer, som kan bruges til at udvikle sprogteknologi. Derfor satte vi os for i vores projekt at lave en BLARK. En BLARK (Basic Language Ressource Kit) indeholder de minimalt nødvendige sproglige ressourcer, som er nødvendige for at udvikle sprogteknologi i et givent sprog (Krauwer, 2003; Maegaard et al., 2006). Idet at BLARK'en har stor betydning for vores arbejde med at udvikle en talegenkender – og forventeligt vil have stor betydning for udviklingen af anden sprogteknologi – vil vi gennemgå de sproglige ressourcer, vi har samlet i vores BLARK, som vi kalder ABLARK (application-based BLARK). Alle delene i ABLARK er blevet samlet, og i skrivende stund arbejder vi mest med at få mere volumen på de enkelte ressourcer. Vi har følgende ressourcer i vores ABLARK: SAMPA, PoS, ordbog, baggrundskorpus, læsemateriale, wav-filer, transskriptionskorpus, værktøjer og dokumentering og vejledninger. Vi vil i det følgende præsentere de enkelte ressourcer.

- SAMPA: Vores internt udviklede maskinlæsbare fonetiske alfabet er inspireret af SAMPA-projektet. Vores SAMPA er kompatibelt med IPA og indeholder de mest almindelige og distinktive foner og diakritika for færøsk (Rischel 1964; Pétur Helgason 2003; Kristján Árnason 2011; Höskuldur Thráinsson et al. 2012; Knooihuizen 2014; Petersen and Adams 2014; Weyhe 2014). Figur 1 viser en del af vores SAMPA.

- PoS: Vores PoS-tagset er inspireret af og kompatibelt med det paneuropæiske PAROLE-tagset (Bilgram and Keson 1998). Et PoS-tagset er et markeringsystem, der markerer hver ordform for sin grammatiske funktion. Figur 2 viser et eksempel på et PoS-tag.
- Ordbog: Vores ordbog indeholder stort set alle funktionsord og uregelmæssige indholdsord og derudover en stor del højfrekvente og mindre frekvente indholdsord. Den indeholder som minimum alle de ord, der optræder i vores transskriptionskorpus. Hver indgang har en udtale, et PoS-tag og et frekvenstal. I skrivende stund har vi ca. 20.000 lemmaer. Figur 2 viser et eksempel på et lemma fra vores ordbog med PoS-tag og udtale.
- Baggrundskorpus: Vi har både et lyd- og et tekstkorpus, som bruges til at træne talegenkenderen. Lydkorpuset kan ikke frigives open source, men dele af tekstkorpuset kan.
- Læsemateriale: Tekster, som vi samler ind og modificerer til vores formål. Teksterne skal bruges til optagesessioner. I den første del af projektet havde vi en ordliste, en fraseliste, en oplæsning af et lukket domæne (fx numerallier), to korte tekster samt en fri samtale (spontantale). Alle disse tekster var konstrueret således, at de skulle eliciterer alle mulige foner i færøsk. I den sidste del af projektet er vi vejet fra denne opsætning, og nu handler det i højere grad om at få mere volumen af optagelser og alle mulige slags tekstgenrer.
- WAV-filer: Lydoptagelser af oplæsning af læsemateriale. I skrivende stund har vi ca. 113 timers optagelse. Alle dialekter er repræsenteret.
- Transskriptionskorpus: Alle optagelser bliver ortografisk transskriberet og får tidskoder, og nogle af dem er blevet fonetisk transskriberet.
- Værktøjer: Sproglige værktøjer og hjælpemidler, som er blevet udviklet i løbet af projektet.
- Dokumentering/vejledninger: Alt vores arbejde er blevet grundigt dokumenteret, og der er lavet vejledninger til vores ordbogsarbejde, PoS-taggingssystem, ortografiske og fonetiske transskriberingskonventioner, optagesessioner osv.

Phoneme group		SAMPA	Length	IPA	Example	Example SAMPA
Consonants	Plosives	p	-	p <sup>h</sup>	pera	p%e:ra
		b	-/+	b	bera, labbi	b%e:ra, l%ab:I
		t	-	t <sup>h</sup>	tala	t%EA:la
		d	-/+	d	dala, koddi	d%EA:la, k%Od:I
		k	-	k <sup>h</sup>	kona	k%o:na
		g	-/+	g	gala, sjagga	g%EA:la, S%ag:a
	Fricatives	f	-/+	f	fara, skuffa	f%EA:ra, sg%Uf:a
		v	-	v	vera	v%e:ra
		s	-/+	s	sigla, kassi	s%i:ja, k%as:I
		S	-/+	ʃ	skip, trýss	S%i:b, tr%UJS:
		z	-/+	ʒ	úrslit, mars	"Yzli:d, m%az:
		h	-	h	húgva	h%Igv:a
	Affricates	tS	-	tʃ <sup>h</sup>	tjúkkur	tS%YHg:Ur
		dZ	-/+	dʒ	geva, ikki	dZ%e:va, "IHd:ZI
	Nasals	m	-/+	m	gloymd, mamma	gl%OJmd, m%am:a
		M	-	ɱ	gloymt	gl%OJMd
		n	-/+	n	máni, tunn	m%OA:nI, t%Un:
		x	-	ɲ	tunt	t%Uxd
		N	-	ɳ	sangur	s%aNgUr
		X	-	ɳ̥	rankur	r%EXgUr
	Laterals	l	-/+	l	lítill, ball	l%UJ:dIl, b%al:
		L	-	ɭ	kalt	k%aLd
	Approximants	j	-	j	ja	j%EA:
w		-	w	mugu	m%u:wU	
r		-/+	ɹ	ringur, fyrr	r%INGUr, f%Ir:	
Semivowels, together with diphthongs	J	-/+	i	feitur, feitt	f%aJ:dUr, f%aJHt:	
	W	-/+	u	tómur	t%OW:mUr. h%aW:n	
	A	+	a	spakur	sb%EA:gUr	

Figure 1. En del af det færøske SAMPA-alfabet. Her ses konsonantinventaret.



ORTO:fiskur	PPOS:NCMSN==IU	PHON:f%lsgUr
ORTO:fisk	PPOS:NCMSA==IU	PHON:f%lsg
ORTO:fiski	PPOS:NCMSD==IU	PHON:f%lSdZl
ORTO:fisks	PPOS:NCMSG==IO	PHON:f%lsgs
ORTO:fiskurin	PPOS:NCMSN==DU	PHON:f%lsgUrn
ORTO:fiskin	PPOS:NCMSA==DU	PHON:f%lSdZln
ORTO:fiskinum	PPOS:NCMSD==DU	PHON:f%lSdZlnUn
ORTO:fisksins	PPOS:NCMSG==DO	PHON:f%lsgslns
ORTO:fiskar	PPOS:NCMP[AN]==IU	PHON:f%lsgar
ORTO:fiskum	PPOS:NCMPD==IU	PHON:f%lsgUn
ORTO:fiska	PPOS:NCMPG==IO	PHON:f%lsga
ORTO:fiskarnir	PPOS:NCMPN==DU	PHON:f%lsgarnlr
ORTO:fiskarnar	PPOS:NCMPA==DU	PHON:f%lsgarnar
ORTO:fiskunum	PPOS:NCMPD==DU	PHON:f%lsgUnUn
ORTO:fiskanna	PPOS:NCMPG==DO	PHON:f%lsgana

Figure 2. Eksempel på et lemma fra vores ordbog. Her ses substantivet *fiskur* 'fisk' med tilhørende PoS-tag og lydskrift. Som eksempel kan vi forklare det første tag *NCMSN==IU*: *N* = noun, *C* = common, *M* = masculine, *S* = singular, *N* = nominative, *I* = indefinite, *U* = Unmarked.

Alle de sproglige ressourcer i vores BLARK fungerer som et økosystem, hvor de forskellige ressourcer er afhængige af hinanden, samtidig som de også vokser sig større og lærer af hinanden. Et eksempel er, at alle ord, der forekommer i vores læsetekster og transskriptioner, også skal være repræsenteret i ordbogen, alle ord i ordbogen skal forsynes med lydskrift med vores SAMPA-tegn og PoS-tag fra vores PoS-tagging-system osv.

Alle ressourcer fra vores BLARK bruger vi til at udvikle en færøsk talegenkender, men ideen med BLARK'en er også, at den skal kunne bruges til udvikling af andre sprogteknologiske værktøjer og sprogforskning generelt. Derfor arbejder vi frem mod, at så mange af vores ressourcer som muligt skal være frit tilgængelige (open source), således at andre kan få nytte af dem og bruge dem til videre udvikling. Derfor bruger vi også kun åbne filformater (txt, csv, rtf, textGrid, wav, flac).

### 3. Status på og udfordringer med færøske sprogteknologiske værktøjer og ressourcer

Af færøske sprogteknologiske værktøjer findes allerede en talesyntese og en stavekontrol. Talesyntesen har eksisteret siden 2005 (Pétur Helgason et al. 2005) og bliver flittigt brugt, især af blinde og ordblinde. Det er en historie om en ildsjæl, der drømte om en talesyntese som hjælpemiddel til de blinde – og som viste sig også at blive meget værdifuld for ordblinde og læsesvage. Det er selvsamme ildsjæl, der har sat gang i talegenkendelsesprojektet. I 2013 blev talesyntesen fornyet med to ny stemmer – en kvinde- og en mandestemme – som fungerede bedre end den første.

Derudover har vi fået en helt ny stavekontrol, hvis beta-version blev lanceret i 2020. Der har eksisteret en stavekontrol i flere år, men nu er der altså

kommet en ny, forbedret version. Den er lavet som et samarbejde mellem det Færøske Universitet og Universitetet i Tromsø (UiT).

Der findes ikke nogen talegenkender endnu, men den er vi som bekendt i gang med at udvikle, og den bliver forhåbentlig færdig i 2022. Derudover er der nogle, der sidder og arbejder med automatisk maskinoversættelse.

Af færøske sproglige ressourcer findes der også lidt af hvert. Der findes en række digitale ordbøger, korpora og PoS-taggere. Dog støder vi på nogle udfordringer i forbindelse med disse ressourcer. For det første er der ofte copyright eller andre adgangsmæssige begrænsninger på dem, så det kan være svært at få fingrene i dem og bruge dem til sprogteknologiske formål. For det næste er flere at disse ressourcer og værktøjer hverken udviklet eller forankret på Færøerne. Det gør det svært at finde ud af, hvad der faktisk findes, hvor det findes, hvem man skal kontakte for at få fat i det osv. Derudover er det et problem, at nogle af disse ressourcer ikke findes i et format, der er kompatibelt med sprogteknologi, dels fordi de er af ældre dato, dels fordi de er lavet med henblik på klassisk sprogforskning og ikke sprogteknologi.

Disse udfordringer med format, copyright og manglende oversigt over ressourcerne forsøger vi at løse med vores BLARK, i og med at den kommer til at være frit tilgængelig, ressourcerne er i åbent format, og alle ressourcerne er samlet samme sted.

#### **4. Sprogbevarelse og domænetab**

Når vi skriver om talegenkendelse, sprogteknologi, ressourcer og udfordringer i et lille sprog som færøsk, så skal vi måske også tale lidt om, hvorfor manglende sproressourcer og sprogteknologiske programmer og værktøjer er et problem. Som i små sprogsamfund flest er sprogbevarelse og styrkelse af sproget en levende ideologi på Færøerne. Færøsk har gennem alle tider været påvirket af andre sprog. De kulturelle og statslige forbindelser til Danmark har resulteret i stor dansk indflydelse på det færøske sprog. Se Petersen (2010) og Hagström (1997) for en grundigere gennemgang af den færøske og danske sprogkontakt. I nyere tid mærker vi også, at engelsk begynder at trænge sig ind i det færøske hverdagsprog, især mellem de unge (Andreasen 2021). Selvom færøsk er det dominerende sprog, er både dansk, engelsk og andre sprog en del af det færøske sprogsamfund. Især visse sproglige domæner fungerer i praksis som flersproglige. Et eksempel på et kulturelt sprogligt domæne påvirket af dansk og engelsk er handel, varer og produkter. Dette er et resultat af, at den største del af færøsk handel er import. Produktnavne er danske og brugervejledninger er engelske eller danske. Som sprogteknologer fokuserer vi mere på de teknologiske sproglige domæner: styresystemer (Windows, iOS), digita-

le assistenter (Siri, Alexa), søgemaskiner (Google, Bing), computerspil, diverse brugerprogrammer, digital kommunikation m.m. Mange færøske sprogbrugere gør brug af disse tjenester og programmer hver dag og er afhængige af dem for at kunne arbejde og kommunikere med hinanden. Dette resulterer i, at færøske sprogbrugere læser meget dansk og engelsk i deres interaktion med teknologi. Og hvis de vil gøre brug af stemmestyringsfunktioner på deres telefon, må det ligeledes gøres på dansk eller engelsk. Det teknologiske er derfor et sprogligt domæne, som færøsk har svært ved at komme ind i. Hvis målet er at bevare det færøske sprog som kommunikationssprog i alle samfundets processer og områder, er det nødvendigt at få færøsk ind på det teknologiske område. Et led i dette arbejde er bl.a. at udvikle taleteknologiske værktøjer, som vi gør i Projekt Ravn. Men der skal mere til end det. Det kræver en målrettet indsats fra hele den sprogpoltiske og sprogteknologiske sektor.

En anden udfordring, som små sprog alle har til fælles, er, at de internationale tech-giganter ikke har interesse i at implementere færøsk i deres produkter. De har ikke den kulturelle interesse i det, og sproget er for småt, til at der er økonomisk interesse. Man kan håbe, at en open source-tilgang til udvikling af materiale og værktøjer kan gøre dette lettere, så vi kan få færøsk ind i alle de store programmer og produkter, som vi alle sammen bruger.

Hvis man som sprogsamfund arbejder frem mod sprogbevarelse som ideologi, så vil det være et naturligt og nødvendigt led i dette arbejde at styrke færøsk sprogteknologi som helhed. Færøsk sprog må ind i flere sproglige domæner, og især de tekniske sproglige domæner er vigtige at fokusere på. Dette vil kræve en organiseret sprogstrategi med konkrete projekter.

## **5. Sprogteknologiens udfordringer på Færøerne og fremtiden**

For at opsummere den færøske sprogteknologiske situation, så er den præget af sprogpoltiske udfordringer. Og når vi siger udfordringer, så er det måske lidt bagvendt, fordi sprogteknologi er faktisk blevet prioriteret de seneste år med vores projekt om talegenkendelse, en stavekontrol og maskinoversættelse, og det er store og vigtige skridt. Men udfordringen er nu at sørge for at vedligeholde og videreudvikle på hele den sprogteknologiske front. Vi er flere, der arbejder med det, og der er flere entusiaster, der gør et stort (og til dels frivilligt) arbejde. Det ville dog styrke området med en form for organisatorisk centralisering af sprogteknologisk virksomhed i færøsk, både af teknisk og sproglig karakter.

Der er altså to områder, der giver udfordringer: finansiering og administration/centralisering af sprogteknologisk virksomhed. Det offentlige har finan-

sieret en stor del af talegenkenderen, og har også delvist finansieret andre sprogteknologiske værktøjer. Udfordringen ligger dog i, at der tidligere ikke har været afsat penge til at videreudvikle og vedligeholde de sprogteknologiske værktøjer, der allerede er udviklet, samt til at udvikle ny sprogteknologi. Og det er selvfølgelig et problem, fordi sproget er i konstant udvikling, og teknologien må nødvendigvis følge med den sproglige udvikling. Dette ser dog ud til at blive finansielt prioriteret de kommende år.

Til sidst er der den udfordring, at i modsætning til mange andre lande, så er der ingen samlende instans på Færøerne inden for sprogteknologi. Der er altså ingen, der samler, vedligeholder og videreudvikler den sprogteknologi, der allerede er. Der findes ikke nogen instans, man kan kontakte og spørge, hvad der findes af sprogteknologi på Færøerne, hvor man kan finde forskellige korpora, eller hvem man skal kontakte for at få fat på talesyntesen osv. Håbet er, at al færøsk sprogteknologi bliver forankret fx i Málráðið (det færøske sprognævn) eller Færøernes Universitet, eller at et sprogteknologisk center bliver udviklet til formålet.

Vi har nu skrevet en del om udfordringer. Men vi vil også understrege, at det går den rigtige vej, og vi er optimistiske. Der er aktivitet på det taleteknologiske og sprogteknologiske område, og udviklingen går i den rigtige retning. Og vi håber, at open source-materiale kan facilitere endnu mere arbejde på området.

Og derudover er vi meget spændte på at byde den nordiske sprogteknologiske konference Nodalida til Færøerne i 2023, hvor vi skal være værter for konferencen. Vi glæder os til at bidrage og forventer, at det vil give os endnu mere skub i den rigtige retning.

## Summary

This article is written on the basis of a talk on speech recognition and language technology in the Faroe Islands at Nordisk Språkmøte 2021. The article presents the ongoing project on Faroese speech recognition and the circumstances regarding the project. A big part of the project consists in developing language resources for language technology. We call this a BLARK (Basic Language Resource Kit), and the BLARK for Faroese is detailed in the article. In short, a BLARK consists of the minimal set of language resources needed for developing language technology. When describing the work on developing speech recognition and language resources for Faroese, a consideration on the status and the challenges of the language technology situation in the Faroe Islands is also needed and is therefore also presented in the article.

## Forfattere

**Sandra Saxov Lamhaug** og **Iben Nyholm Debess** er begge to cand.mag. i lingvistik. De er ansat hos fonden Talutøkni i projekt Ravn, som udvikler den første færøske talegenkender.



Sandra Saxov Lamhaug



Iben Nyholm Debess

## Referencer

- Adams, Jonathan & Hjalmar P. Petersen, 2014: *A Language Course for Beginners*, 3. udg. Tórshavn: Stiðin
- Andreasen, Rakul Skaale, 2021: *Útlitini fyri einum málskipti: Fara vit at tosa enskt?* Tórshavn: Fróðskapur, Setursrit 12.
- Árnason, see Kristján Árnason.
- Bilgram, Thomas and Keson, Britt, 1998: The Construction of a Tagged Danish Corpus. *Proceedings of the 11th Nordic Conference of Computational Linguistics*, NODALIDA 1998, s. 129-139.
- Hagström, Björn, 1997: Faroese-Danish. I: Nelde, Peter H., Stry, Zdenek, Goebl, Hans & Wolck, Wolfgang: *Kontaktlingvistik / Contact Linguistics, Part 2*. Berlin/Boston: De Gruyter, 1043-1049.
- Helgason, see Pétur Helgason.
- Höskuldur Thráinsson, Hjalmar P. Petersen, Jógvan í Lon Jacobsen & Zakaris Hansen, 2012: *Faroese, an overview and reference grammar*, 2. udg. Tórshavn: Fróðskapur and Reykjavík: Linguistic Institute, University of Iceland
- Knouihuizen, Remco, 2014: Variation in Faroese and the development of a spoken standard: In search of corpus evidence. *Nordic Journal of Linguistics*, 37(1): 87-105.
- Krauwer, Steven, 2003: The Basic Language Resource Kit (BLARK) as the First Milestone for the Language Resources Roadmap. I: *Proceedings of the International Workshop "Speech and Computer"*, SPECOM 2003, Moscow, Russia.

- Kristján Árnason, 2011: *The Phonology of Icelandic and Faroese*. Oxford University Press, Oxford, UK.
- Maegaard, Bente, Steven Krauwer, Khalid Choukri & Lise Damsgaard Jørgensen, 2006: The BLARK concept and BLARK for Arabic. I: *Proceedings of the 5th International Conference on Language Resources and Evaluation*, s. 773- 778.
- Petersen, Hjalmar P., 2010: *The Dynamics of Faroese-Danish Language Contact*. Heidelberg: Universitätsverlag Winter
- Pétur Helgason, 2003: Faroese Preaspiration. I: *Proceedings of the 15th International Congress of Phonetic Sciences*, s. 2517-2520. Universidad Autónoma de Barcelona, Barcelona, Spain.
- Pétur Helgason, Sjúrdur Gullbein & Karin Kass, 2005: Færøsk talesyntese: Rapport marts 2005. I: Henrik Holmboe (red.): *Nordisk sprogteknologi 2005 - Årbog for Nordisk Sprogteknologisk Forskningsprogram 2000-2004*, s. 51-58.
- Rischel, Jørgen, 1964: Toward the Phonetic description of Faroese vowels. *Fróðskaparrit*, 13:99-113.
- Thráinsson, see Höskuldur Thráinsson.
- Weyhe, Eivind, 2014: Variatión av i og u í herðingarveikari støðu í føroyskum. *Fróðskaparrit*, 61:116-136.

## **Nøgleord**

talegenkendelse, sprogteknologi, Færøerne, BLARK, ressourcer

# Samisk språkteknologi i 2021

*Sjur Nørstebø Moshagen  
Institutt for språk og kultur  
UiT Noregs arktiske universitet*

*Artikkelen gjev eit oversyn over samisk språkteknologi i 2021, i lag med eit stutt samandrag av historia, og går deretter inn på nokre av utfordringane i arbeidet med samisk språkteknologi og språkteknologi for minoritetar meir allment. Utfordringane blir tydeleggjorde med handfaste døme. Artikkelen blir avslutta med å peika på nokre vegar framover for å rydda dei viktigaste hindera av vegen.*

## 1. Oversyn

Arbeidet med samisk språkteknologi starta tidleg på 1990-talet, då professor Pekka Sammallahti og professor Kimmo Koskenniemi laga eit utkast til ein nordsamisk morfologisk analysator. Kring år 2000 vart arbeidet deira gjeve vidare til UiT Noregs arktiske universitet (UiT) ved professor Trond Trosterud. I etterkant av dette vart Giellatekno-gruppa<sup>1</sup> ved UiT oppretta for å arbeida vidare med samisk språkteknologi.

I 2004 vart Divvun-gruppa<sup>2</sup> starta ved det norske Sametinget, og arbeidet med å utvikla retteprogram for nord- og lulesamisk byrja for alvor i 2005, i nært samarbeid med Giellatekno-gruppa og med ein felles kodebase. Verktøya vart lanserte i 2007, og i 2008 starta arbeidet med å utvikla tilsvarande verktøy for sørsamisk. Desse vart lanserte i 2010. I 2011 vart Divvun-gruppa flytta frå det norske Sametinget til UiT og samtidig gjort permanent, i staden for å vera prosjektfinansiert. Frå då av har UiT vore senteret for forskning og utvikling av språkteknologi for samiske språk og andre urfolks- og minoritetsspråk – som til dømes færøysk og grønlandsk – med eit miljø på i snitt ti personar. Over tid har det òg vorte utvikla ein avansert og språkuavhengig infrastruktur som gjer det lett å starta arbeid med eit nytt språk og utvikla dei mest grunnleggjande verktøya ganske raskt.

---

1 <https://giellatekno.uit.no/index.nob.html>

2 <https://divvun.no/>

## 2. Kva finst i dag?

Over desse ikring 20 åra har det vorte utvikla ein god del verktøy. Det finst no ein infrastruktur som gjer det enkelt å definera utforminga av **tastatur**, både for datamaskiner og mobile einingar. Med denne infrastrukturen har UiT-miljøet bygt datamaskintastatur for dei fleste samiske språka, og utforminga er gjort i samarbeid med Giellagáldu<sup>3</sup>, det allnordiske, samiske språknormeringsorganet. Det er dessutan bygt mobiltastatur for både samiske og andre språk.

**Retteprogramma** var og er ein sentral del av mykje av det arbeidet som er og framleis blir gjort ved UiT. Det starta med stavekontrollar for nord- og lulesamisk som vart bygt ut med automatisk orddeling ganske raskt, og deretter støtte for sørsamisk. Dei siste åtte åra er kjernen i retteprogramma tufta på teknologi frå Helsingfors universitet (Lindén m.fl. 2009). Verktøykassa har i seinare tid vorte vidare utvida med ein avansert grammatikkontroll basert på teknologi frå Syddansk universitet (Karlsson 1995; Bick & Didriksen 2015).

For minoritetsspråk er tilgang til gode **ordbøker** heilt sentralt, og UiT-miljøet har utvikla og tilbode elektroniske ordbøker i over ti år. Dei elektroniske ordbøkene er dei mest populære ressursane som finst i Tromsø, og har ein samla søkjefrekvens på mange tusen søk i døgnet. Det er to hovudportalar for ordbøkene, med ulik profil og ulike brukarar: NDS<sup>4</sup> og satni.org<sup>5</sup>.

Språkteknologien som er utvikla ved UiT, blir òg brukt i nettbaserte **språklæringsprogram**<sup>6</sup> for fleire samiske og andre språk. Programma analyserer det studentane svarar, og gjev tilbakemeldingar basert på analysen.

I samarbeid med Apertium<sup>7</sup> er det utvikla **maskinomsetjing** for fleire par av språk som ein kan omsetja mellom (språkpar). Det einaste språkparet som er offentleg tilgjengeleg, er nordsamisk–norsk, men det blir aktivt jobba med fleire andre, mellom anna nordsamisk–lulesamisk, i samarbeid med den nord-samiske dagsavisa Ávvir<sup>8</sup>.

Sidan 2015 har det eksistert eit **tekst-til-tale**-system for nordsamisk, utvikla i samarbeid mellom Divvun, Sametinget og Acapela<sup>9</sup>, og sidan slutten av 2020 er det eit nytt prosjekt på gang for å gjera det same for lulesamisk. Det meste

3 <https://www.giella.org>

4 <https://dicts.uit.no>

5 <http://satni.org>

6 <https://oahpa.no>

7 <https://www.apertium.org>

8 <https://www.avvir.no>

9 <https://www.acapela-group.com/>



av arbeidet i det nye prosjektet blir gjort ved UiT, og all programkode skal vera open kjeldekode. Det er òg starta eksperiment med **talegjenkjenning**.

Språkteknologimiljøet ved UiT har i samarbeid med det norske Sametinget samla inn eit stort **samisk tekstkorpus**. Tekstane i korpuset er analyserte, tilgjengelege og søkbare i ein nettapp.<sup>10</sup> Det gjer det lett for alle å sjå til dømes korleis eit spesifikt ord blir brukt i ulike tekstar og samanhengar.

Alle verktøya som er utvikla ved Sametinget og UiT, er grammatikkbaserte og altså ikkje bygde på statistiske eller andre maskinlæringsbaserte metodar. Det er ein av hovudgrunnane til at det er mogleg å laga alle desse verktøya, sjølv for språk med få eller ingen elektroniske ressursar. Trass i at teknologi, infrastruktur og røynsle finst, er det likevel fleire nordiske minoritetsspråk og urfolksspråk som framleis har få verktøy for å kunna brukast digitalt. Det gjeld framfor alt romanispråka og dei minste samiske språka.

Tromsø-miljøet samarbeider med Uleåborg universitet, Anarâškielâ servi og Meän akateemi, i tillegg til institusjonar i Russland og Nord-Amerika, for å tilby infrastruktur og kunnskap knytt til utvikling av språkteknologi for minoritetar og urfolksspråk. Per i dag har vi grammatiske modellar for ca. 130 ulike språk og tastaturdefinisjonar for 50 språk, alle lagra i kjeldekodehandteringssystemet GitHub<sup>11</sup>. Nokre av modellane og tastatura er nærmast som fragment å rekna, medan andre er fullt utvikla og i dagleg bruk av tusentals språkbrukarar.

### 3. utfordringar i 2021

Alle nordiske land har visjonar om å digitalisera samfunnet. Landa har òg språkpolitiske føringar. Den styrande tanken er at alle digitale tenester skal vera tilgjengelege for alle innbyggjarar, på innbyggjaren sitt eige språk, og minoritetsspråka skal vera samfunnsberande i sine egne regionar, slik det er nedfelt i den nordiske språkdeklarasjonen<sup>12</sup>. Dette vil ikkje fungera om ikkje alle delane av den digitale infrastrukturen er lagde til rette for fleirspråklegheit og er tilpassa kvart enkelt språk.

Det er mykje som enno haltar, også for dei statsberande majoritetsspråka. Men det haltar mykje meir for minoritetsspråka, til eit punkt der det knapt går å nytta språka digitalt i visse samanhengar. I det følgjande vil eg gje nokre døme på slike bakevjer i den digitale infrastrukturen.

---

10 <https://gtweb.uit.no/korp>

11 <https://github.com/giellalt>

12 <https://www.norden.org/no/declaration/sprakdeklarasjonen>

### 3.1 Udefinerte språk

For at eit språk skal fungera i den digitale verda, må språket vera mogleg å identifisera. Det betyr at operativsystem og dataprogram veit at det eksisterer eit slikt språk, og kan slå opp kva for verktøy som finst for det spesifikke språket. For mange av språka vi arbeider med, er ikkje dette tilfellet.



Figur 1. Slik vil ein ikkje sjå språket sitt.

Konsekvensane av at eit språk ikkje er definert, er fleire. Brukarane vil berre sjå ein språkkode i staden for det riktige namnet på språket. Det gjer at dei ikkje kjenner att språket sitt, og verktøya blir såleis utilgjengelege. På Windows blir det uråd å nytta korrekturprogram, og det blir umogleg å nytta operativsystemtenester for språket, sjølv om ein installerer dei komponentane som trengst. I tillegg sender det eit kraftig signal til språkbrukarane om at språket deira ikkje er mykje verd.

Den enklaste løysinga på dette er å overføra ansvaret for alle delstandardane i ISO 639 til ein felles organisasjon (dei er i dag spreidde på fleire institusjonar), til dømes Unicode<sup>13</sup>-organisasjonen, og deretter behandla heile ISO 639-familien på same måten som ein behandlar Unicode-standarden: Heile standarden er lagt inn i alle operativsystem, og han blir oppdatert med jamne mellomrom.

### 3.2. Tastatur

For norsk-, svensk-, dansk-, islandsk- og finskspråklege, og fleire til, er det sjølv sagt at datamaskiner, mobiltelefonar og nettbrett har tastatur for det språket brukarane nyttar. Dette er ikkje sjølv sagt for minoritetsspråka, og dei som har, har det ikkje takk vera Microsoft, Google eller Apple – med unntak for færøysk og nordsamisk.

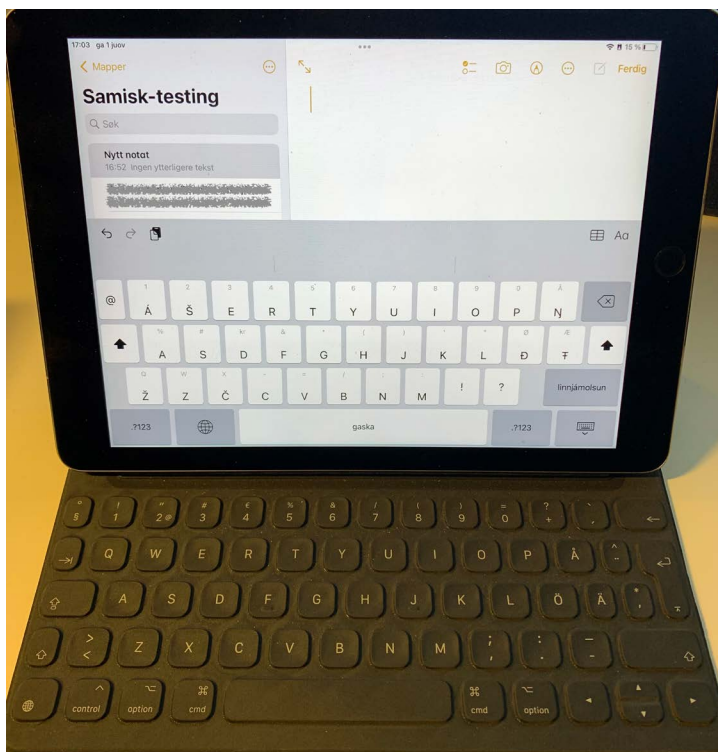
Språkteknologimiljøet ved UiT har jobba med tastaturløysingar i mange år

13 <https://home.unicode.org/>

og har no eit system som gjer det enkelt å laga tastatur for nye språk. Med dei verktøya som finst, har det vorte laga tastatur for dei fleste samiske språka, både mobiltastatur og datamaskintastatur. Likevel er det ein del tilfelle der ting ikkje fungerer.

Mange skular har teke i bruk nettbrett eller Chrome-bøker for elevane sine og bruker desse som hovuddatamaskin i undervisinga. Nettbretta kan sjølvsagt nytta tastatura vi har laga, men både for Android-nettbrett og for Apple sine iPad-ar er dei fysiske tastatura blokkerte for oss – vi får ikkje lov til å skriva tastaturdefinisjonar som kan bli brukte i lag med dei fysiske tastatura. Det betyr at samiske elevar heile tida må byta mellom det fysiske tastaturet (som er på norsk, svensk eller finsk) og det samiske tastaturet på skjermen. Det går – men det er ikkje slik ein vil arbeida ein heil skuledag.

Eit anna døme er Chrome-bøkene frå Google. I desse finst det ikkje innebygd støtte for eit einaste samisk språk. Det har vi gjort noko med, men tastatura er framleis variantar av majoritetsspråka. Det er Google sine rammer som gjer at vi må gå fram på denne måten for at tastatura skal fungera.



Figur 2. Og aldri skal dei to tastatura møtast. Nordsamisk på skjermen, svensk/finsk fysisk tastatur.

iOS-tastatura vi lagar, vil alltid vera bleike kopiar av Apple sine. Apple legg føringar og hindringar i vegen for oss som gjer at vi, sjølv med svært dyktige programmerarar, aldri kan tilby same funksjonalitet som Apple sjølv. Eit trivielt døme er emoji-tastaturet: På alle Apple sine eigne tastatur, som dei for norsk, svensk osv., ligg det ein emoji-knapp til venstre for mellomromstasten. Det gjer det lett å leggja inn ein emoji i teksten, ein funksjonalitet som så godt som alle nyttar no for tida. Men Apple hindrar oss i å utvikla denne funksjonaliteten for samiske brukarar. På grunn av restriksjonane Apple legg på oss, kan vi difor ikkje tilby ei like saumlaus og god brukaroppleving som det Apple kan, for samiske brukarar. Anten må vi klara oss utan – samiske brukarar må i staden trykkja på globusen, byta tastatur, skriva emoji og så byta tilbake – eller vi må laga vårt *eige* emoji-tastatur og halda det oppdatert heile tida. Det har vi vurdert som altfor kostbart.

Hindringane som teknologigigantane legg i vegen for oss som arbeider med språkteknologi for små språk, gjev oss ekstra kostnader og ekstra arbeid. Det blir nær umogleg å kunna tilby det folk ventar seg. Vi betalar slik ein slags «minoritetsspråksskatt» fordi dei store programvareprodusentane ikkje gjer ein skikkeleg jobb i høve til minoritetsspråka. Slik burde det ikkje vera i 2021.

### 3.3. Retteprogram

Noko av det fyrste som vart planlagt og laga då arbeidet med samisk språkteknologi byrja, var ein stavekontroll. Det er eit enkelt og etter kvart sjølv sagt verktøy for å fanga opp vanlege skrivefeil, og verktøyet er svært viktig i eit språksamfunn der mange er usikre på rettskrivinga. Dei fyrste stavekontrollane for nord- og lulesamisk kom i 2007 og har vore viktige hjelpemiddel sidan då (Antonsen & Trosterud 2020).

I mange år var den vanlegaste framgangsmåten at ein installerte eit retteprogram på si eiga datamaskin eller på ein server som IT-folka styrte med, slik at stavekontrollen fanst lokalt. Men dei siste åra har det kome ei endring, ein ny måte å laga og bruka kontorprogram på: i nettlesaren. Og pluteleg låg ikkje stavekontrollen på eiga datamaskin eller på serveren til IT-folka – han låg på serveren til Google, Microsoft eller andre IT-kjempar. Og dei vil ikkje ha verktøy og program som dei ikkje har laga sjølve. Med eitt slag var samisk språkteknologi tilbake til tida før 2007.

Dette gjeld all programvare over nettet, såkalla SaaS<sup>14</sup>, og rårkar minoritetsspråk mykje hardare enn majoritetsspråk. Med SaaS er det ikkje lenger mogleg

---

14 SaaS = Software as a Service

å installera egne tillegg, slik som ein stavekontroll, og om det er, er både funksjonaliteten og brukargrensesnittet kraftig redusert.

Det er ikkje umogleg å få til ei fungerande løysing som tek vare på personvern og datatryggleik. Til dømes kunne dei store IT-selskapa ha egne, sertifiserte koplingspunkt som andre skulle kunna nytta etter avtale, samtidig som brukarane skulle bli informerte om kva organisasjon retteprogramma kjem frå. Men ingen av dei store selskapa har så langt gjort nokre forsøk i ei slik retning.

## **4. utfordringar framover**

UiT-miljøet ved Divvun- og Giellatekno-gruppene har vist at det er mogleg å byggja grammatikkbasert språkteknologi for minoritetsspråk og for språk med små eller ingen elektroniske ressursar. Det går an å laga funksjonelle og nyttige verktøy for alle språksamfunn i ein elektronisk kvardag. Dette er dessutan heilt naudsynt for at språka skal halda fram med å vera bruksspråk i sine egne samfunn. Det kjem ikkje av seg sjølv, og arbeidet som har vorte gjort, er eit resultat av ei målretta satsing av det norske Sametinget og det norske Kommunal- og distriktsdepartementet (KDD) over mange år. Utan innsatsen deira ville vi ikkje hatt dei verktøya vi trass alt har i dag.

Samtidig ser vi at digitaliseringa berre går raskare og raskare, og frå andre urfolkssamfunn og minoritetar rundt om i verda veit vi at trykket mot dei små språka berre aukar. Mange språk kjem til å gå or bruk og døy ut dette hundreåret.

Det neste store steget innanfor språkteknologi ser vi allereie: digitale assistentar som ein kan prata med og få munnlege svar tilbake frå. Dei er enno ganske enkle, men fungerer likevel godt for dei domena dei kan handtera. Hovudproblemet er at dei berre finst for dei største og økonomisk mest interessante språka.

Og slik kjem det i all hovudsak til å bli verande. Ein kan ikkje venta at Google, Microsoft, Apple eller andre tek på seg å laga språkteknologi for 7000 språk, heller ikkje for dei omkring 3500 språka som framleis vil vera i dagleg bruk om hundre år (Crystal 2000). Derimot veit vi at grupper som språkteknologimiljøet ved UiT vil arbeida for å gje verktøy til utsette språksamfunn.

Utfordringane knytte til slikt arbeid er både teknologiske og språklege.

### **4.1. Teknologiske utfordringar**

Språkteknologi, særleg avansert språkteknologi som maskinomsetjing og digitale assistentar, er viktige drivkrefter i den digitale økonomien og er ikkje noko som dei store selskapa lett opnar opp. Like fullt er det det dei må. Dei må gjera det mogleg for andre partar enn dei sjølve å utvikla og levera teknologiske

løysingar på dei språka dei sjølve ikkje tek ansvar for. Det er mest synleg med taleteknologi – skal ein kunna prata samisk med bilen sin i framtida, må nokon laga eit samisk system for talegjenkjenning og talesyntese – men gjeld heile den språklege infrastrukturen, frå menytekstar til Siri. Det må vera mogleg for kvart språksamfunn å sjølve ta styringa over teknologien som gjeld deira eige språk. Om ikkje Google, Apple, Microsoft eller dei andre store vil ta ansvar for eit språk – og det er mange språk dei ikkje kan ta ansvar for – så skal dei ikkje samtidig hindra andre i å gjera det.

Det betyr i praksis at alle delar av operativsystema og tenestene dei leverer som gjeld naturlege språk, må opnast opp. Dei treng ikkje opna opp den underliggjande teknologien, men dei må gjera det mogleg for andre å plugga inn sin eigen teknologi for språk som dei sjølve ikkje har teknologi for. Alle operativsystem skil i dag mellom språkspesifikke delar og språkuavhengige komponentar. Detaljane varierer sjølvsagt, men skiljet er der. Det betyr at grunnlaget for å gjera dei språkspesifikke delane tilgjengelege eller opne finst. Det vil krevja ein del arbeid, men det er eit naudsynt arbeid om alle språksamfunn skal kunna bli ein del av det digitale samfunnet på like vilkår. Når eit slikt skilje er gjennomført, dei språkspesifikke delane er opne og andre har byrja å laga slike komponentar for nye språk, kan alle dei store systemprodusentane distribuera språkpakker frå tredjepartar gjennom sine eigne programvarebutikkar.

## 4.2. Språklege utfordringar

Dei språklege utfordringane handlar om ressursar og språkutvikling. Teknologien vi nyttar ved UiT, er grammatikkbasert, og det trengst folk som kan grammatikken til språka og kan skriva han om til noko som datamaskinene forstår. Det trengst òg arbeid med terminologi og leksikografi, og det trengst arbeid med skriftspråksnormering. Dette arbeidet er ein viktig del av det å ta tilbake språka og gjera dei samfunnsberande, og arbeidet er heilt naudsynt om ein skal oppfylla intensjonen bak den nordiske språkdeklarasjonen.

## 5. Oppsummering

Eg har i denne artikkelen gjeve eit oversyn over det arbeidet som har vorte gjort for samisk og andre minoritets- og urfolksspråk ved UiT. Gjennom arbeidet vårt har vi vist at det er mogleg å laga gode verktøy for språk med få ressursar og kompleks grammatikk. Samtidig har eg peikt på utfordringar knytte til det å få tilgang til verktøya i ein del samanhengar og ulike hindringar som dei store teknologiselskapa legg i vegen for dei små språksamfunna. Eg har samtidig prøvd å peika på moglege vegar framover. Målet må vera at kvart samfunn sjølv skal kunna ta ansvar for sin eigen språkteknologi, i samarbeid

med partar som dei sjølve kan velja. Berre slik kan vi byggja ei digital framtid som inkluderer alle.

## Summary

In this article I give an overview over what has been achieved by the Divvun and Giellatekno groups at *UiT The Arctic University of Norway* over the 20 years since the work started, both in terms of concrete products and tools, but also in infrastructure development enabling new language communities to enter the digital train. I then discuss some concrete issues we have met concerning the IT giants' lack of attention regarding the consequences of their actions towards minority and indigenous language communities. I end the article by pointing out how these issues could be solved in general, by opening up the field of language technology components to third parties and let them offer language packs for languages not supported by the tech giants, through the app stores of the platforms. These steps are necessary to minimise language death as much as possible.

## Forfattar

**Sjur Nørstebø Moshagen** har arbeidd med språkteknologi i både industri og akademia i ikring 30 år og har vore leiar for Divvun-gruppa ved UiT sidan starten i 2004.



## Litteratur

Antonsen, Lene & Trond Trosterud, 2020: Med et tastetrykk. Bruk av digitale ressurser for samiske språk. / Boallobevavddi bokte. Sámegeielaid digitála resurssaid geavaheapmi. I: *Sámi logut muitalit 13. Čielggaduvvon sámi statistibkka 2019/Samiske tall forteller 13. Kommentert samisk statistikk 2020. Sámi allaskuvla.*

Bick, Eckhard & Tino Didriksen, 2015: [CG-3 - Beyond Classical Constraint Grammar](#). I: Beáta Megyesi (red.): *Proceedings of NODALIDA 2015, May 11-13, 2015, Vilnius, Lithuania*. s. 31–39. Linköping: LiU Electronic Press.

- Crystal, David, 2000: *Language Death*. Cambridge University Press, Cambridge.
- Karlsson, F., A. Voutilainen, J. Heikkilä, & A. Anttila, 1995: *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*, Mouton de Gruyter, Berlin.
- Lindén, Krister Mikkilä Silfverberg & Tommi Pirinen, 2009: HFST Tools for Morphology – An Efficient Open-Source Package for Construction of Morphological Analyzers. I: Mahlow, Cerstin & Michael Piotrowski: *State of the Art in Computational Morphology*, s 28–47. Springer Berlin Heidelberg.

## **Nøkkelord**

språkteknologi, minoritetsspråk, språkpolitikk



# Ny språklitteratur 2021

## Danmark

*Marianne Rathje*

**Andersen, Margrethe Heidemann og Eva Skafte Jensen (red.): Med fornøden agtelse. Festskrift til Jørgen Schack på 60-årsdagen 14. maj 2021.** Dansk Sprognævns skrifter 52. 2021. 312 s.

I denne antologi, der er et festskrift til seniorforsker i Dansk Sprognævn Jørgen Schack, kan du læse 21 artikler, der spænder over mange emner, fx: Christian Becker-Christensen: R-problemer, Dorthe Duncker: Performativer i dansk, Jørgen Nørby Jensen: *Bjørnetjeneste* – et falsk pendulord?, Henrik Jørgensen: *Man* og *en* – paradigme eller ej?, Sabine Kirchmeier: Ikke en kina-persons chance? Om sammensætninger med *-mand*, *-kvinde* og *-person*, Ole Ravnholt: Trap ned fra startkommaerne, Hanne Ruus: Liv i det centrale ordforråd – kerneord med 40 års interval.

**Andersen, Nina Møller: Ordet i livet og ordet i kunsten. Bachtins sprogbrugsbegreber.** UPress. 2021. 168 s.

Michail Bachtin tilhører litteratur- og kulturforskningens mest citerede tænkere i nyere tid. Men der har længe været forvirring om brugen af hans grundbegreber. Det rådes der dog bod på med denne bog. Et af formålene med bogen er at klargøre Bachtins begreber, gøre dem entydige og anvendelige til sprogbrugsanalyser.

**Becker-Christensen, Christian: Dansk Retskrivning. Bogstav – Lyd – Bogstav. En grafemisk beskrivelse af dansk ortografi.** Syddansk Universitetsforlag. 2021. 408 s.

Bogen handler om ordenes ortografi set i forhold til det talte sprog, om retskrivningens struktur og funktionalitet og om dens forudsætninger og vilkår. Relationerne mellem bogstav og lyd kortlægges i et omfang svarende til ordforrådet i Retskrivningsordbogen (2012) udgivet af Dansk Sprognævn. Den standardsproglige udtale, som står for den ene side i bogstav-lyd-relateringen,

bliver der redegjort for fonetisk og fonologisk. Bogens enkelte hovedafsnit følges op med uddybende omtale af begreber, terminologi og analyser refererende til relevant litteratur.

**Bjerring-Hansen, Jens, Simon Skovgaard Boeck og Eva Skafte Jensen (red.): Nogle betænkninger om dansk sprog og litteratur. Festskrift til Marita Akhøj Nielsen.** Syddansk Universitetsforlag. 2021. 650 s.

Dette festskrift til Marita Akhøj Nielsen består af 31 bidrag skrevet af forskere og kollegaer. Af bidrag til bogen kan fx nævnes: Lene Schøsler: Danismer i Leonora Christinas fransk, Henrik Jørgensen: Konsekvens i renæssancens retskrivning?, Hans Basbøl: Radikal ortografi og/eller lydskrift? Nogle eksempler fra Rasmus Rask og J.H. Bredsdorff.

**Dansk Sproghistorie V: Dansk i samspil.** Red. af Ebba Hjorth, Henrik Galberg Jacobsen, Bent Jørgensen, Birgitte Jacobsen, Merete Korvenius Jørgensen og Laurids Kristian Fahl. Det Danske Sprog- og Litteraturselskab og Aarhus Universitetsforlag. 2021. 375 s.

Formålet med *Dansk Sproghistorie* er at skabe og viderebringe ny erkendelse om og indsigt i det danske sprogs historie fra de første runeindskrifter til dagens sprog og dermed bidrage til den sproglige refleksion i Danmark, både blandt læg og lærd.

Bind 5 handler om dansks store indflydelse på de lokale sprog i de lande som gennem historien har været en del af Danmark. Den situation beskrives sammen med danske udvandreres fortsatte brug af deres modersmål i de nye hjemlande. Særlige kapitler beskriver brugen af dansk i EU og brugen af dansk som kommunikationsmiddel med andre nordboer i dag og den internordiske sprogforståelse. Endelig gennemgås hvilken rolle undervisning i dansk sprog spiller ved udenlandske universiteter mv. Gennem historien har andre sprog end dansk i perioder været talt i og haft indflydelse i Danmark. Tidligere eksempler er latin, tysk, fransk, romani, jiddish. Nu er det nyere indvandrersprog som serbokroatisk, tyrkisk, kurdisk, urdu og arabisk. Politisk, økonomisk og kulturelt samarbejde med andre lande har kunnet og kan betinge brugen af andre sprog i Danmark. Det gælder russisk og ikke mindst engelsk. De øvrige nordiske sprog og grønlandsk høres også i Danmark.

**Danske Studier 2021.** Red. af Simon Skovgaard Boeck og Henrik Blicher under medvirken af Carl-Emil Prehn Andersen. Universitets-Jubilæets danske Samfund nr. 603. 2021. 282 s.

Af bindets indhold er der i denne sammenhæng grund til at fremhæve Eva

Skaftø Jensen: Et system i stadig bevægelse. Pronominalkasus i dansk, Bolette Sandford Pedersen, Sanni Nimb og Sussi Olsen: Dansk betydningsinventar i et datalingvistisk perspektiv. Katrine Kehlet Bechsgaard: Personnavne og forenelighed. Hvordan passer familiens navne sammen?

**Dansknoter.** Red. af Birgitte Darger m.fl. 2021. Nr. 1-3. 62 s. + 70 s. + 66 s.

*Dansknoter* er et fagtidsskrift for undervisere i de gymnasiale uddannelser samt et kulturtidsskrift hvor danskfaget og danskfagets placering i offentligheden er til diskussion og refleksion. Af artikler i dette års numre kan fremhæves, fx Lene Rotne: Stil- og genrebrud i gymnasiestile – resultaterne af et forskningsprojekt (nr. 1), Marianne Rathje og Nicoline Østergaard-Nielsen: Sproglig korrekthed som danskundervisningens forsømte stedbarn? (nr. 3), Ditte Zachariassen: Du er nødt til at forstå det yani, ellers du tror bare det er sjusk (nr. 3), Jesper Tinggaard Svendsen, Anna Vibeke Lindø, Alexandra Emilie Møller Holsting og Marianne Rathje: Det kan godt være der er problemer med kommaerne og retstavningen, men ... (nr. 3).

**Greve, Linda: Ord.** Aarhus Universitetsforlag. Tænkepauser 86. 2021. 60 s.

I begyndelsen var ordet. Påstod nogen få årtier e.Kr. Men et ord er ikke bare et ord fra tidernes morgen til dommedag. Vores vigtigste værktøj til kommunikation er snarere under evig forvandling. Vi opfinder nye ord, ændrer de gamles betydning eller lader dem gå i glemmebogen. Sprogets mindste betydningsbærende enheder opstår ikke ud af intet, men i et syndigt rod af private tanker, kropslige erfaringer og omgivelsernes påvirkning.

**Hansen, Marianne Haugaard: Sproglige generationsforskelle på sociale medier: En sproglig etnografisk undersøgelse af unge og deres forældres sprogbrug online.** Københavnerstudier i tosprogethed bind 81, Københavns Universitets Humanistiske Fakultet. 2021. 92 s.

I dette nummer af Københavnerstudier undersøges det hvordan to generationer kommunikerer sprogligt på sociale medier, og hvilke forestillinger de har om den anden generations måde at kommunikere på.

**Holmen, Anne og Helene Thise: Flere sprog i læringen. Translanguaging i praksis.** Samfundslitteratur. 2021. 179 s.

I denne bog præsenterer forfatterne det internationale begreb translanguaging og giver en lang række eksempler på, hvordan du som lærer kan arbejde med elevers translanguaging i praksis og i forskellige fag. Gennem bogen får du indsigt i translanguagingbegrebet set i et dansk perspektiv, værktøjer til at

tilrettelægge tværsproglige forløb i mange fag, inspiration til konkrete aktiviteter og didaktiske greb til direkte brug i din daglige praksis, eksempler på translanguaging i praksis, redskaber til at planlægge, gennemføre og evaluere undervisning inspireret af translanguaging og inspiration til at udvikle skolens, teams og egen praksis gennem refleksionsspørgsmål og udviklingsaktiviteter

**Kjæræk, Laila og Ditte Boeg Thomsen (red.): Børns sprogtilegnelse. Sproglig udvikling hos danske børn i alderen 0-6 år.** Frydenlund Academic. 2021. 376 s.

Børns sprogtilegnelse er en forskningsbaseret grundbog i sproglig udvikling hos børn i alderen 0-6 år. Med grundlag i sprogvidenskab, psykologi og pædagogik giver bogen en samlet indføring i børns sprogtilegnelse og præsenterer børns sproglige udvikling som en integreret del af deres kognitive og sociale udvikling. Bogen giver en nuanceret forståelse af processer og faktorer i både typiske og atypiske udviklingsforløb og viser, hvordan indsigterne fra den nyeste danske og internationale forskning kan anvendes i den daglige pædagogiske praksis.

Bogen indeholder bl.a. kapitlerne "fonologisk udvikling", "ordtilegnelse", "grammatisk udvikling", "kommunikativ udvikling og pragmatiske færdigheder", "to- og flersprogede børn" og "sprogudvikling hos børn med udviklingsmæssige sprogforstyrrelser".

**Tore Kristiansen og Anne Holmen (red.): Sprogs status i Rigsfællesskabet år 2031.** Københavnerstudier i tosprogethed, Studier i Parallelsproglighed bind C13, Københavns Universitets Humanistiske Fakultet. 2021. 196 s.

Denne bog er det fjerde bind i en serie der er kommet med et nyt bind hvert tiende år siden 1991 med danske sprogforskeres syn på hvordan sprogsituationen vil udvikle

sig ti år frem i tid. Af relevante artikler i antologien kan nævnes: Jógvan í Lon Jacobsen: Færøsk i et fremtidsperspektiv, Camilla Kleemann-Andersen: Den evigt nærværende koloniale fortid, Jacob Thøgersen: "At bevare" og "at udvikle" det danske sprog: DR's sprogpoltiske forpligtelser og afrapporteringer af sprogpoltisk arbejde, Lian Malai Madsen og Andreas Candefors Stæhr: Sociale medier – standardsvækkelse og sproglig liberalisering?, Nicolai Pharao: Dansk udtale i 2031 – året hvor vi har lært at tale ordentligt?, Tore Kristiansen og Marie Maegaard: "Rigsdansk og dialekterne" i fremtidsperspektiv,

**Mål & Mæle.** 42. årgang. Red. af Ditte Boeg Thomsen, Simon Poulsen, Ken Farø og Thomas Olander. 2021. Nr. 1-3. 32 s.

Foruden brevkassen *Sprogligheder* med spørgsmål om sprog indeholder *Mål & Mæle* artikler om sprog. Af disse numres indhold er der grund til at fremhæve Lene Rotne: Fra Facebookstil til skolestil – uden slinger i valsen (nr. 1), Louise Fischer-Nielsen: Det store i det små. Hvordan oversætter man Bibelen til nudansk? (nr. 1), Mikkel Wallentin: Sproget i hjernen. Neurolingvistik før og nu (nr. 3).

**Nielsen, Bodil: Grammatik i funktion. På opdagelse i sproget.** Samfundslitteratur. 2021. 258 s.

*Grammatik i funktion* viser, hvordan læreren kan undervise i grammatik tæt knyttet til elevernes arbejde med de tekster, de læser og skriver. Der er gennemgang af masser af teksteksempler og forslag til opgaver, man kan arbejde med i undervisningen. Dertil enkle forklaringer på grammatikken. Bogen indeholder bl.a. kapitler om didaktiske principper bag undervisning i grammatik i funktion, ordklassers brug og betydning, sammensatte ord, brug af morfemer og medbetydninger og billedbrug. Bogen henvender sig til lærerstuderende, lærere og læsevejledere.

**Ny forskning i grammatik 28.** Red. af Anna Sofie Hartling, Michael Nguyen, Peter Juul Nielsen og Jørgen Schack. Dansk Sprognævn 2021. 194 s.

*Ny forskning i grammatik* giver læseren et indblik i den nyeste nationale sprogforskning inden for dansk og en lang række fremmedsprog. Af bindets artikler kan nævnes: Tanya Karoli Christensen: Trusler, tilskyndelse – eller “bare” hadtale? Placering af agenten i udsagn om fremtidig skade, Katalin Fenyvesi: Prøv at se her! – grammatikalisering af verbet *prøve* i dansk, Viggo Bank Jensen: Finsk kasus i dansk sprogforskning. Fra Rask til Hjelmlev og Ohrt, Michael Nguyen: Hvornår er præpositionsstranding i dansk umuligt?

**NyS. Nydanske Sprogstudier 59.** Red. af Elisabeth Muth Andersen, Tanya Karoli Christensen, Ken Ramshøj Christensen, Jan Heegård, Michael Nguyen, Mads Poulsen, Line Sandst, Jacob Thøgersen. Udgivet af Dansk Sprognævn 2021. 207 s.

Bind 59 indeholder 5 artikler: Simon Borchmann: Omhed i dansk – umarkeret og markeret, Bjarne Simmelkjær Sandgaard Hansen: Forgrundsinformationsnominativ eller kasusløshed. En analyse af kasussystemerne i den gamelskånske tekst *Sjælens Trøst*, Laura Bang Lindegaard og Line Sandst: Værdiladede betingelsessætninger i argumenterende tv-debat, Marianne Haugaard Hansen og Andreas Candefors Stæhr: Sproglige generationsforskelle på de sociale medier, Alexandra Holsting, Marianne Rathje, Jesper Tinggaard Svend-

sen og Anna Vibeke Lindø: Hvad retter de? Nyindskrevne studerendes revision af universitetsopgaver.

Bind 59 indeholder også Ole Tøgebys anmeldelse af Dansk Sproghistorie, bind 4: Dansk i brug.

**NyS. Nydanske Sprogstudier 60.** Red. af Elisabeth Muth Andersen, Ken Ramshøj Christensen, Jan Heegård, Michael Nguyen, Mads Poulsen, Line Sandst, Jacob Thøgersen. Udgivet af Dansk Sprognævn 2021. 250 s.

Bind 60 indeholder 6 artikler: Peter Juel Henriksen: Glemte ord - En undersøgelse af H.C. Ørstedes nyord og deres plads i nudansk, Ida Flørke og Elisabeth Engberg-Pedersen: Køn i korpus. Semantisk asymmetri i brugen af ordene *kvinde* og *mand*, *pige* og *dreng* i danske avistekster, Martin Hauerberg Olsen og Mads Poulsen: Analyse af krævende syntaks i danskfagets tekster, Holger Juul og Jan Heegård Petersen: Om analyse af stavefejl. Årsager, korrespondens-regler og udtalevariation, Ditte Zachariassen og Tine Lykkegaard Nielsen: 'Eow', 'lak', 'yani' og 'dig' – fire nyere interjektionaler som viser holdninger om vidensforhold, Michael Nguyen: Mulige NP-placeringer ved retningsadverbium + præposition.

Bind 60 indeholder også Merete Birkelunds anmeldelse af Nina Møller Andersens bog Bakhtins sprogbrugsbegreber. Ordet i livet og ordet i kunsten og Hartmut Haberlands anmeldelse af Lita Lundquists Humorsocialisering. Hvorfor er danskerne ikke så sjove som de selv tror?

**Nyt fra Sprognævnet.** Dansk Sprognævn. Publiceret på <https://dsn.dk/nyt/nyt-fra-sproгнаevnet>. 2021, nr. 1-3, 14 s. + 21 s. + 24 s.

Numrene indeholder et udvalg af sproglige spørgsmål til Dansk Sprognævn, bogomtaler og artikler om sproglige forhold, fx Anna Sofie Hartling: Bag om ordene: Samtykke og frivillighed (nr. 1), Niels G. van Leeuwen: Fire måder hvorpå engelsk har præget dansk grammatik (nr. 2), Marianne Haugegaard Hansen: Sproglige generationsforskelle på sociale medier (nr. 3), Nicoline Østergaard-Nielsen og Marianne Rathje: Holdninger til sproglig korrekthed i gymnasiet (nr. 3).

**Ord & Sag 41.** Aarhus Universitet 2021. 72 s.

*Ord & Sag* udgives af Peter Skautrup Centret for Jysk Dialektforskning ved Aarhus Universitet. Nummeret indeholder videnskabelige såvel som populærvidenskabelige artikler om jysk sprog, fx Mette-Marie Møller Svendsen: Jysk Ordbog og dens brugere – en undersøgelse, Viggo Sørensen: *Ligge/lægge* i jysk regi.

**Petersen, Jan Heegård, Holger Juul, Nicolai Pharao og Marie Maegaard: Udtalt. En introduktion til dansk fonetik.** Samfundslitteratur. 2021. 209 s.

*Udtalt* er en øvelsesbaseret indføring i dansk fonetik. Gennem 14 kapitler klæder bogen dig på til at kunne arbejde med og forholde dig til de grundlæggende emner i dansk fonetik.

Bogen kan bruges som grundbog på alle uddannelser der kræver indsigt i dansk talesprog, fx læreruddannelsen, audiologopædi, retorik, dansk som andetsprog og dansk og lingvistik på universiteterne.

**Rask. International Journal of Language and Communication 53.** University of Southern Denmark, Department of Language and Communication 2021. 117 s.

Fra nummeret kan nævnes: Uwe Kjær Nissen: *A flea in the ear and a louse over the liver*. Idiomatic expressions of insects and other small animals located inside the human body.

**Schjørring, Maren Aarup og Ellen Rønhede Hansen: Grammatiknøglen. Almen sprogforståelse.** Dansk lærerforeningens Forlag. 2021. 118 s.

*Grammatiknøglen* er en elevhenvendt bog til almen sprogforståelse i gymnasiet. Bogen giver eleverne en solid indføring i både grammatik, semantik og pragmatik.

**Sprog & Samfund. Nyt fra Modersmål-Selskabet.** 2021, nr. 1, 2, 3, 4. 20 s. + 24 s. + 20 s. + 20 s.

Foruden anmeldelser af sprog litteratur og omtaler og referater af sproglige arrangementer indeholder numrene kortere artikler om forskellige sproglige forhold, fx Martin Lund: Helsætningsstrukturen som grammatisk gøgeunge (nr. 3).

**Sprogforum. Tidsskrift for sprog- og kulturpædagogik 72.** Roskilde Universitetsforlag 2021. 134 s.

Nr. 72 er et temanummer om *Digitale ressourcer* og indeholder bidrag af bl.a. Sarah Bro Trasmundi og Cecilie Lemmeke Mandsen: Unge bruger emojis til at udtrykke sig på nye måder i digital kommunikation, Marnie Lai: At lære med de digitale ressourcer i gymnasiets sprogfag.

**Sprogforum. Tidsskrift for sprog- og kulturpædagogik 73.** Roskilde Universitetsforlag 2021. 100 s.

Nr. 73 er et temanummer om *Brobygning* og indeholder bidrag af bl.a.

Jan Lindschouw, Dorte Albrechtsen, Lars Behnke, Joyce Kling, Sanne Larsen, Natalia Morollón Martí, Cilia Reebirk og Anna Sandberg: Akademisk literacy i engelsk, fransk, spansk og tysk – et brobygningsperspektiv, Kristine Bundgaard, Lotte Dam og Rikke Hartmann Haugaard: Brobygning til sproguddannelser på universitetet. Fem praksisanbefalinger.

**Sprogkræsen. Nyt fra Den Danske Sprogkreds.** 4. årgang, 2021, nr. 10, 11, 12, 13. 32 s. + 32 s. + 28 s. + 32 s.

Numrene indeholder omtaler og referater af sproglige arrangementer, interviews med sprogpersoner, anmeldelser af sprog litteratur og sprogarrangementer samt kortere artikler om forskellige sproglige forhold, fx Lise Bostrup: Er Dansk Sprognævns forskning på afveje? En kritisk analyse af Sprognævnets vurdering af det danske sprogs tilstand i dagens Danmark (nr. 11), Michael Falkendorff: Anders And-sprog (nr. 11), Lise Bostrup: Udtrykker mennesker med depression sig anderledes end andre? (nr. 12), Lise Bostrup: Plutte? Kender I ikke til det, derinde i Københavnsstrup? Regionalsprog på Lolland, Falster, Sydsjælland og de omkringliggende øer (nr. 13), Lise Bostrup: Hvor mange bogstaver er ø, ö, oe og oe? Historien om de danske ø'er (nr. 13).

**Sufi, Ali og Tobias Cadin: GADE/DANSK ORDBOG.** Grønningen 1. 2021. 245 s.

*GADE/DANSK ORDBOG* er din guide til ghettodansk. Lær at forstå, hvad de rapper om i radioen, og mød en subkultur, der taler et helt andet dansk end det, du kan slå op i en ordbog. Ordbogen beskriver mere end 450 ord med oprindelse, bøjninger og eksempler.

**18. Møde om Udforskningen af Dansk Sprog, Aarhus Universitet. 8.-9. oktober 2020.** Red. af Yonathan Goldshtein, Inger Schoonderbeek Hansen og Tina Thode Hougaard. 2021. 542 s.

Siden 1986 har sprogforskere fra ind- og udland hvert andet år mødtes i Aarhus for at sætte fokus på udforskningen af det danske sprog. På det 18. møde deltog 88 forskere (digitalt) med i alt 47 foredrag. En række af dem er trykt i den foreliggende rapport. Af rapportens artikler kan nævnes Margrethe Heidemann Andersen: Fucking nice. Engelske lån i Ex on the beach, Yonathan Goldshtein: Stødets naturlige historie, Inger Schoonderbeek Hansen og Yonathan Goldshtein: Er København (virkelig) eneste sproglige normcenter?, Peter Juel Henriksen: H.C. Ørsted som leverandør til dansk sprogteknologi, Kathrine Thisted Petersen: Det er lidt sådan noget jeg tænker – den



vage formulerings status i uformel dansk sprogbrug sammenlignet med tysk, Søren Sandager Sørensen: Tvungne ja/nej-svar som interviewformat, Ditte Zachariassen og Tine Lykkegaard: 'Lak' og 'eow' – nye(re) interjektioner i aarhusiansk.

## Finland

*Maria Fremer*

**Ahlholm, Aku, 2021: Vargens röst. Rovdjursdiskursen i finsk, finlands-svensk och svensk press från 1990-talet till 2010-talet.** Jyväskylä: Jyväskylä universitet. 240 s.

I avhandlingen undersöks nyhetstexter om vargen. Vargnyheter är både konfliktnyheter och miljönyheter. Objektet, vargen, är tyst och journalister strukturerar sina nyhetstexter med hjälp av andra röster. Syftet är att åstadkomma en modell som hjälper till att identifiera de olika diskursiva representationerna av rösterna och undersöka nyhetstexternas dialog mellan människor med olika världsbilder.

Publikationen finns tillgänglig på: <http://urn.fi/URN:ISBN:978-951-39-8513-4>

**Biström, Anna & Johanna Wassholm (red.), 2021: Historiska och litteraturhistoriska studier 96.** Helsingfors: Svenska litteratursällskapet i Finland. 288 s.

Temat för HLS 96 är teknologi och känslor, och motsättningarna mellan dem. Sorg är i fokus i artiklar om stormaktstidens gravdikter och den sociala elitens brev på 1800-talet. Adrian Pereras romaner analyseras mot bakgrund av frågor om känslor och språklig (in)kompetens. Det organiska livet i en mekanistisk tid diskuteras utifrån Solveig von Schoultz lyrik, medan kärleksbegreppet fokuseras i en jämförelse mellan Tito Colliander och Dostojevskij. Därtill studeras Henrik Tikkanen "inter artes". I numret finns även essäer om resandets känslor och teknologi, tillfällesdikt och receptionen av K.A. Tavaststjernas roman Hårda tider. Dessutom ingår Ulla-Lena Lundbergs festtal från SLS årshögtid.

Publikationen finns tillgänglig på <https://hls.journal.fi/>.

**Eloranta, Ville & Lotta Jalava, 2021: Sana sanasta. Suomen kielen jäljillä.** [Ord om ord. På spaning efter finska språket.] Helsingfors: Tammi. 285 s.

Boken berättar hurdana betydelser, strukturer och berättelser som ligger bakom många vardagliga ord, namn och uttryck. Finskans förflutna är fullt av möten mellan folkslag och märkliga sammanträffanden. Språket avslöjar hur man har levt och vad som har varit viktigt och värt att säga. Språket får ständigt nya ord och strukturer som lever vid sidan av de gamla eller tar över efter dem. Boken presenterar de nyaste rönen om etymologierna och berättar om varför språket över huvud taget förändras.

**Forsberg, Ulla-Maija, 2021: Stadin slangin etymologinen sanakirja.** [Etymologisk ordbok över Helsingforsslang.] Helsingfors: Gaudeamus. 287 s.

Slangen i Helsingfors är en fascinerande blandning av olika vardagsspråk från början av 1900-talet, med inslag från musikens och filmernas drömvärld. Ofta behövs djupgående kunskaper om ungdomskulturen i Helsingfors och om stadens historia när man vill ha reda på ursprunget till orden, som kan ha lånats in från svenska, ryska, tyska, italienska eller finska. Ordboken innehåller cirka 2 600 olika uppslagsord som framför allt representerar den äldre slangen, från 1890-talet till 1960-talet. Ordboken är avsedd för alla som är intresserade av finska, språk i allmänhet och etymologi, samt – naturligtvis – för dem som använder och hör Helsingforsslang.

**Häggman, Johan, 2021: Strategier för minoritetsspråk. Jämförelser av relevans för svenskan i Finland.** Helsingfors: Tankesmedjan Magma. 106 s.

Svenskan har en stark juridisk ställning i Finland, men räcker det för att säkra en fortsatt tvåspråkighet i landet? Ska vi nöja oss med att ha ”världens bästa språklagstiftning”, eller ska vi aktivt främja det svenska språket genom språkplanering? En genomgång av språkstrategier i Europa och Kanada visar flera exempel på hur antalet minoritetsspråkstalare blivit fler med hjälp av aktiva åtgärder. Bland annat satsar Kanada på att genom invandring höja antalet och andelen fransktalande kanadensare. Exempelen i rapporten understryker utbildningens viktiga roll för språket. Samtidigt bör kultur- och medieutbudet understödjas för att bidra till att göra språket tillgängligt och attraktivt för både svensk- och annanspråkiga.

Publikationen finns tillgänglig på <https://magma.fi/svenskfinland-har-mycket-att-lara-av-andra-sprakminoriteter/>.

**Jagne-Soreau, Maïmouna, 2021: Postinvandringslitteratur i Norden. Den litterära gestaltningen av icke-vita födda och uppvuxna i Norden.** Helsingfors: Helsingfors universitet. 194 s.

Postinvandringslitteraturen kännetecknas av återkommande teman, som identitet ur ett mellanförskapsperspektiv, det vill säga en varken-eller-dialektik vad gäller känsla av tillhörighet. Därtill återkommer temat generationskonflikt mycket starkt, och bekräftar nödvändigheten av att förstå skillnaden i upplevelsen av en generation som har invandrat, och en som inte gjort det. Temana rasism, rasifiering och vithet som överlägsen norm är också ständigt närvarande och tyder på ett akut behov av att diskutera dessa ämnen ytterligare i samhället. I avhandlingen analyseras ett tjugotal verk, bland annat av Maria

Navarro Skaranger från Norge, Yahya Hassan från Danmark, Erik Lundin från Sverige och Adrian Perera från Finland.

Publikationen finns tillgänglig på: <http://urn.fi/URN:ISBN:978-951-51-7715-5>

**Juva, Kersti, 2021: Tolkienin tulkkina. Tarina Sormusten herran suomentamisesta.** [Tolkiens tolk. Berättelsen om att översätta Sagan om ringen.] Helsingfors: SKS Kirjat. 232 s.

Kersti Juva fick som 23-åring i uppdrag att översätta en av århundradets viktigaste romaner till finska. Efter trilogin Sagan om ringen fortsatte hon, och har nu översatt nästan hela J.R. Tolkiens produktion. I den här boken återvänder Juva till början av sin långa karriär som Tolkienöversättare. Hon berättar bland annat om hur romanfigurernas sätt att tala och de många ort- och personnamnen i böckerna fick sina finska motsvarigheter. Hur kommer det sig till exempel att orchernas språk låter som Helsingforsslang? Och hurdan är egentligen översättarens kärlek till det finska språket?

**Kaipainen, Suvi & Annamari Saure, 2021: Äimän käkenä. Äidinkielen puuhakirja aikuisille.** [Språkpysselbok för vuxna.] Helsingfors: SKS Kirjat. 158 s.

Är du en språkpolis? Rättar du skrivfel i din kollegas e-postmeddelande och på paketet med frukostflingor? Eller känns kommatering och semikolon som högre matematik för dig? Det här är en glad och humoristisk övningsbok för alla vuxna som är intresserade av språket. Finskan är inte världens svåraste språk, men särskrivningar orsakar ofta oavsiktlig komik, och ett vårdslöst utplacerat kommatecken kan ändra hela meningens betydelse. Boken fokuserar på sådana språkfenomen som ofta orsakar problem.

**Kairos Future (konsultföretag), 2021: Det svenska i Finland. Vadan och varthän.** Helsingfors: Tankesmedjan Magma. 74 s.

Denna rapport ger ett faktabaserat utifrånperspektiv på det svenska i Finland. Den analyserar förändringar i omvärlden och diskuterar vilka konsekvenser de kan ha för svenskans ställning i Finland. Fokus ligger på demografi, kultur, identitet och skola. Enligt rapporten kommer tvåspråkighet i praktiken bli en norm bland svenskspråkiga år 2040. På grund av nya språkliga minoriteter och engelskans starka frammarsch kan det i framtiden bli allt svårare att sälja in svenskan som en individuell rättighet för en krympande grupp enspråkigt svenska. Som en direkt följd av utvecklingen får den svenska skolan en allt större betydelse som upprätthållare av såväl det svenska språket som av en svensk identitet för unga svensk- och tvåspråkiga.

Publikationen finns tillgänglig på <https://magma.fi/det-svenska-i-finland-ett-utifranperspektiv/>

**Katajamäki, Heli, Mona Enell-Nilsson, Hannele Kauppinen-Räisänen, Liisa Kääntä & Harri Salovaara (red.), 2021: Workplace Communication IV.** VAKKI Publications 13. Vasa: Föreningen för kommunikationsforskning rf. 169 s.

Workplace Communication IV innehåller nio refereegranskade artiklar på finska, svenska, tyska och engelska. De behandlar arbetslivskommunikation ur tre olika perspektiv: För det första tematiseras företagskommunikation ur ett hållbarhets-, transparens- och marknadsföringsperspektiv. För det andra undersöks utvecklandet av kompetenser i arbetslivskommunikation både i ett universitetssammanhang och i en företagskontext. För det tredje granskas experters roll i utvecklingen av fackspråksutbildning.

Publikationen finns tillgänglig på: <https://vakki.net/index.php/2021/12/21/workplace-communication-iv/>

**Kautonen, Maria & Mikko Kuronen, 2021: Uttalsinläring med fokus på svenska.** Helsingfors: Svenska litteratursällskapet i Finland. 216 s.

Uttal är en central del av muntlig färdighet, men trots det finns det i dag rätt lite forskningsbaserad kunskap om hur man lär sig uttal och hur man undervisar i uttal. Vilka faktorer påverkar uttalsinläringen och hur går den till? Boken presenterar olika perspektiv på svenska som inlärningsmål, och speglar dem mot empiriska resultat och teoretiska modeller som fått stöd i senare forskning. Den diskuterar också undervisningsmetoder i praktiken och lyfter fram pedagogiska perspektiv på inläringen.

I boken granskas särskilt inläring av svenska i Finland, men författarna tar upp både finlandssvenska och sverigesvenska.

Publikationen finns tillgänglig på <https://www.sls.fi/sv/utgivning/uttalsinlarning-med-fokus-pa-svenska>.

**Kepsu, Kaisa & Lia Markelin, 2021: Den svenskspråkiga ungdomsbarmetern 2020. Del 2: språk, medier och hobbyer.** Helsingfors: Tanke-smedjan Magma. 47 s.

Svenskregistrerade unga använder mycket svenska i sin vardag och tar relativt ofta del av finlandssvenskt medieinnehåll. De flesta, 60 procent, har en svenskspråkig identitet. De unga anger i snitt att omkring tre fjärdedelar av deras kommunikation i hemmet och med vänner sker på svenska och ca 15 procent på finska. Även ungdomar som identifierar sig som tvåspråkiga använder mer svenska än finska hemma, med vänner och i sociala medier.

Publikationen finns tillgänglig på <https://magma.fi/svenskan-ar-stark-i-de-finlandssvenska-ungdomarnas-liv/>

**Kielikello. Kielenhuollon tiedotuslehti.** [Språkklockan. Finska språkvårdens informationsblad.] Helsingfors: Institutet för de inhemska språken.

Kielikello utkommer fyra gånger om året.

Årets första nummer har temat termer, ord och ordböcker, bl.a. termer som beskriver ett språks ställning i samhället, såsom modersmål, hemspråk, officiellt språk.

Det andra numret behandlar bl.a. en bibelöversättning med nutida ord och uttryck: Nya testamentet i elektronisk utgåva för mobila enheter. En artikel handlar om orden i äldre Helsingforsslang. Svenskans inverkan på slangen har varit betydande och det finns också spår av svenskt fikonspråk.

Temat för årets tredje nummer är naturen och språket. I tidskriften granskas bl.a. nyord i anslutning till klimatförändringen.

Temat för det fjärde numret är språkpolitik. Artiklarna handlar om statsrådets nationalspråksstrategi, om olika alternativ till språkstrategier för flerspråkiga arbetsplatser, och om faktorer som främjar och försvårar invandrades språkinläring.

Publikationen finns tillgänglig på <https://www.kielikello.fi/>.

**Kolehmainen, Leena, Aino Liira & Kirsi-Maria Nummila (red.), 2021: Kieliä ja kohtaamisia Turun historiassa. Näkökulmia monikielisyyteen.** [Språk och möten i Åbos historia. Perspektiv på flerspråkighet.] Helsingfors: Suomalaisen Kirjallisuuden Seura. 360 s.

Finlands äldsta stad har en rik historia som sträcker sig från medeltiden till i dag. Åbo har alltid varit en internationell, mångkulturell och flerspråkig stad. Verket tecknar en bild av Åbos språkliga landskap under olika tidsperioder och visar hur överskridandet av språk- och kulturgränser alltid har varit en fast del av stadslivet. Verket närmar sig flerspråkighet från olika synvinklar: språkvetenskap, historia, arkeologi, litteratur och kulturarvsforskning. Med olika metoder och skriftliga material, föremål, arkeologiska fynd och dialektinspelningar åskådliggör den hur olika språk kopplas till vissa situationer och användningsändamål.

**Lahdenperä, Hanna, 2021: ”Det finns miljoner sätt att vara”. Om Monika Fagerholms Diva och den filosofiska läsningen.** Helsingfors: Helsingfors universitet. 225 s.

I avhandlingen identifieras tre centrala begrepp i Diva: språklighet, kropps-

lighet och rumslighet. De utgör grunden för en filosofisk läsning där begreppen fungerar som motiv i romanen och som en tematisk filosofisk diskussion om berättelsens, litteraturens och representationernas roll för protagonisten.

Publikationen finns tillgänglig på: <http://urn.fi/URN:ISBN:978-951-51-7289-1>

**Leden, Laura, 2021: Adaption av flickskap. Normbekräftande och normbrytande i flickböcker översatta från engelska till svenska och finska 1945–1965.** Helsingfors: Helsingfors universitet. 360 s.

Flickboken handlar om och förhåller sig till att vara flicka, att göra flickskap i en viss tid och på en viss plats. Flickskap är både anpassning till och uppror mot samhällets normer eller förväntningar. Översättningar av flickböcker påverkas av normer för flickskap, normer för berättande och översättningsnormer. Adapterades eller anpassades gestaltningen av flickskap när flickböcker översattes från engelska till svenska och finska, och i så fall hur? Studien söker svar genom en kombination av översättningsvetenskapens, litteraturvetenskapens, språkvetenskapens och arkivforskningens metoder.

Publikationen finns tillgänglig på <http://urn.fi/URN:ISBN:978-951-51-7320-1>

**Lehti-Eklund, Hanna, Jannika Lassus & Caroline Sandström (red.), 2021: Folkmålsstudier 59.** Helsingfors: Föreningen för nordisk filologi. 159 s.

Artiklarna i det aktuella numret av Folkmålsstudier uppvisar tematiskt, teoretiskt och metodiskt en stor bredd. Ett språkhistoriskt perspektiv finns i Marko Lamberg, Sanna Skärlands, Hans Landqvists och Nina Pilkes artiklar även om tidsperioderna och temana skiljer sig från varandra. Saija Tamminen-Parre behandlar språkideologier och attityder i sin text. Maria Kautonen och Mikko Kuronen fokuserar på bedömning och språkinläring i sin artikel som tar upp fonetiska drag i L2-tal. Slutligen ingår en fallstudie av Pirjo Kukkonen och Eeva-Liisa Nyqvist som analyserar översättning från norska till finska av en fantasytrilogi.

Den elektroniska utgåvan är fritt tillgänglig på webben med ett års fördröjning på <https://journal.fi/folkmalsstudier>.

**Pajunen, Anneli & Mari Honko (red.), 2021: Suomen kielen hallinta ja sen kehitys. Peruskoululaiset ja nuoret aikuiset.** [Utvecklingen av färdigheterna i finska hos grundskolelever och unga vuxna.] Helsingfors: Suomalaisen Kirjallisuuden Seura. 493 s.

Boken handlar om finländska grundskoleelevers intuitiva kunskaper om språkriktighet. I verket utreds nivån på språkfärdigheterna, på vilket sätt den

varierar enligt ålder och kön, samt hur den förändras från en åldersgrupp till en annan. Det omfattande empiriska materialet möjliggör för första gången en täckande jämförelse mellan olika åldersgrupper. Den är till nytta både i undervisning och i planering av undervisning, eftersom ett tillräckligt ordförråd och språkfärdighet är en förutsättning för läs- och skrivförmågan.

**Peltola, Rea, Anni Jääskeläinen & Katariina Harjunpää (red.), 2021: Kieli ja eläin. Vuorovaikutusta ja kielioppia monilajisissa yhteisöissä.** [Språk och djur. Interaktion och grammatik mellan människor och djur.] Helsingfors: Suomalaisen Kirjallisuuden Seura. 446 s.

I artiklarna beskrivs de mångfasetterade sätt på vilka människor talar till djur och om djur. Hur svarar du när en hund skäller eller en papegoja hälsar? Hur går det till när en människa och en hund tillsammans leker med en boll? Vad ropar man för att få korna att komma hem? Hur talar riksdagsledamöter från olika partier om djur? Kan vi använda personliga pronomen om djur? Allt det här får man svar på genom att studera förhållandet mellan människor och andra djur ur en språkvetenskaplig synvinkel.

**Pentikäinen, Johanna & Kimmo Svinhufvud, 2021: Riittävän hyvää kirjoittamista. Opas kouluun ja koulunjälkeiseen elämään.** [Skriva tillräckligt bra. En guide för skolan och livet efter skolan.] Helsingfors: Art House. 216 s.

Denna bok är en praktisk och sporrande guide för gymnasieelever och studerande samt modersmållärare och andra som undervisar i skrivande. Boken närmar sig skrivfärdigheten som ett kontinuum från skolan till studentexamen och vidare in i studie- och arbetslivet. Den tar upp många olika aspekter på skrivande, såsom olika texttyper, hur man hittar sin egen röst, och skrivande som process. I boken berättar nyutexaminerade studenter, lärare och föreläsare för olika yrken vilka utmaningar de har upplevt som skribenter och ger tips om hur man hanterar skrivstress.

**Riitamaa, Tomi, 2021: Isolerat och övergivet nationsfragment, hänsynslöst ihjältigt? Studier i den finlandssvenska litteraturens position och predikament i Sverige.** Helsingfors: Helsingfors universitet. 232 s.

Hyllad eller styvmoderligt behandlad – hur tas den finlandssvenska litteraturen egentligen emot i Sverige? I denna litteratursociologiska doktorsavhandling undersöks bland annat den finlandssvenska litteraturens spridning i Sverige, hur kritiker skriver om finlandssvenska böcker i de stora svens-



ka dagstidningarna och hur den finlandssvenska litteraturen behandlas i den svenska litteraturhistorien.

Publikationen finns tillgänglig på: <http://urn.fi/URN:ISBN:978-951-51-7496-3>

**Sadeniemi, Matti, 1951–1961. Nykysuomen sanakirja 1–6.** [Nufinsk ordbok 1–6.] Helsingfors: WSOY. Elektronisk faksimilupplaga 2021, Helsingfors: Institutet för de inhemska språken.

Nufinsk ordbok var den första omfattande ordboken över finskans centrala ordförråd. Den innehåller över 200 000 sökord som representerar språket under 1900-talets första decennier, men också en hel del äldre och dialektalt material. Pdf-utgåvan möjliggör fritextsökning i materialet.

Publikationen finns tillgänglig på: [https://www.kotus.fi/sanakirjat/muut\\_sanakirjat/nykysuomen\\_sanakirja\\_%281951-1961%29](https://www.kotus.fi/sanakirjat/muut_sanakirjat/nykysuomen_sanakirja_%281951-1961%29)

**Saxell, Jani, 2021: Tanssii kirjainten kanssa. Inspiraatiota tämän päivän tarinankertojille.** [Dansar med bokstäver. Inspiration för berättare i vår tid.] Helsingfors: Art House. 296 s.

Dansar med bokstäver är en hängivet personlig guide till kreativt skrivande. Den utgår från grunderna: persongalleri, tidsbeskrivning, strukturella lösningar och olika berättarröster. Dessutom behandlas återberättande av sagor, myter och klassiker, realistiska och fantastiska stadsbeskrivningar, brev- och dagboksromaner och deras motsvarigheter på webben. De rikliga praktiska exemplen öppnar dagens litterära världar.

**Språkbruk.** Helsingfors: Institutet för de inhemska språken.

Språkbruk utkommer sedan 2021 inte längre på papper, utan enbart elektroniskt på tidskriftens egen webbplats. Artiklarna publiceras fortlöpande under hela året. I Språkbruk publiceras populärvetenskapliga artiklar om språk i alla dess former, intervjuer och recensioner. I och med övergången till ett heldigitalt format infördes en ny artikelkategori, Synpunkten, där opinionsartiklar publiceras. Under året har fyra artiklar i kategorin publicerats, med olika teman: ordböcker, klarspråk, språklig variation och språkteknologi. En annan ny artikelkategori är Fem frågor, där den intervjuade besvarar frågor från redaktionen. I Krönikan publiceras mer personliga betraktelser. Under året publicerades sammanlagt 60 artiklar.

Publikationen finns tillgänglig på <https://www.sprakbruk.fi/>.

## Färöarna

*Ragnar Sigrunarson*

**Andreasen, Rakul Skaale, 2021: Útlitini fyri einum málskifti.** Fara vit at tosa enskt? Kanningar av hugburðinum hjá føroyingum til føroyskt, enskt, føroyskt-enskt kotuskifti og til danskt. Setursrit 12. Fróðskapur. Faroe University Press. Tórshavn. 119 s.

Bogen er en undersøgelse af unge færingers holdning til færøsk, dansk, engelsk og færøsk-engelsk kodeskifte. Rakul Skaale Andreasen undersøger, hvordan den sproglige fremtid ser ud på Færøerne med henblik på et muligt sprogskifte fra færøsk til engelsk.

**Angantýsson, Ásgrímur, 2021: The relative order of sentential adverbs in Icelandic and Faroese.** Fróðskaparrit 67, s. 81-95.

This paper discusses the relative order of four types of sentential adverbs in Insular Scandinavian. Data from two judgment tasks show that the relative orders of adverbs that follow Cinque's (1999) hierarchy receive more positive judgments than orders that violate this hierarchy in both languages, but the contrasts are much weaker than expected. That Icelandic and Faroese behave in very similar ways with respect to adverb placement is expected, given all the syntactic similarities between the two languages.

**Jacobsen, Jógvan í Lon, 2021: Ein søga um onk, eymkan og ank.** Fróði 1/2021 27. árg., s. 34-38.

Artiklen undersøger ordene onkur, eymkan og ank med hensyn til fælles etymologi og skiftende betydning.

**Jacobsen, Jógvan í Lon, 2021: Færøerne.** Red.: Ebba Hjort. Dansk sproghistorie bind 5. Dansk i samspil. Aarhus Universitetsforlag og Det Danske Sprog- og Litteraturselskab, s. 15-39.

Da de norske vikinger drog på deres rejser vestover, kom deres sprog i brug på de steder uden for Norge, hvor de slog sig ned, bl.a. på Færøerne i 800-tallet. De første århundreder efter bosætningen på Færøerne, det såkaldte *landnam*, benyttede landnamsmændene og deres efterkommere det norrøne sprog i tale og skrift, som de havde taget med sig fra det vestlige Norge. Men i takt med at det norske politiske herredømme i senmiddelalderen svækkedes, mistede den færøske variant af det oprindelige, vestnorske

eller norrøne sprog efterhånden sin rolle som administrations- og kirkesprog til fordel for dansk.

**Jacobsen, Jógvan í Lon, 2021: Færøsk i et fremtidsperspektiv.** Red.: Tore Kristiansen og Anna Holmen. Sprogs status i Rigsfællesskabet 2031. Københavnerstudier i tosprogethed. Studier i parallelsproglighed. Københavns Universitet. Humanistisk Fakultet, bind C 13, s. 15-27.

I denne artikel fokuseres der på følgende temaer: dialekt, grammatik, holdning til dialekt, holdning til purisme og fagsprog. Artiklen er en statusrapport med et forsigtigt blik ind i fremtiden.

**Jacobsen, Jógvan í Lon, 2021: Føroysk purisma. Føroysk orð ella orð í føroyskum?** Fróðskapur. Faroe University Press. Tórshavn. 264 s.

Bogen beskriver og behandler færøsk purisme i en historisk så vel som en samtidig kontekst. En teoretisk del drejer sig om sproglige variationer, holdninger, normer og konvergens. Bogen består også af en nutidig del, der analyserer låneord og neologismer. Analysen er baseret på undersøgelsen "Moderne importord i språka i Norden" (2002-2012).

**Jacobsen, Jógvan í Lon, 2021: Purisma og hjáveiða.** Red.: Raket Helmsdal, Sanna Andrassardóttir Dahl og Sanna Klein. Varðin 88. bind, s. 26-32.

Artiklen argumenterer for, at purismen på Færøerne har været årsag til, at det færøske ordforråd er vokset betydeligt, at purismen derved har faciliteret en væsentlig sprogvariation på øerne, og at den dertil har forårsaget det betydelige skel mellem talesprog og skriftsprog.

**Petersen, Hjalmar P., 2021: Føroysk mállæra 3.** Ljóðlæra. Nám. Tórshavn. 96 s.

*Føroysk mállæra* er en moderne færøsk grammatik egnet til gymnasie-, universitets- og seminarienniveau foruden til andre med interesse for det færøske sprog. Dette tredje bind af Hjalmar Petersens nutidsfærøske grammatik drejer sig om fonetik.

**Petersen, Hjalmar P., 2021: Uppávegin – kanska tað kemur frá uppávegin = 'tung'.** Frøði 1/2021 27. årg., s. 32-33.

Artiklen argumenterer for en alternativ etymologi for det færøske adjektiv *uppávegin* 'gravid'. Hjalmar P. Petersens hypotese er, at *uppávegin* er en udvikling af frasen *vigað er upp á bana* 'der er vejret på hende'.

**Weyhe, Eivind, 2021: Vatnanøvn í Føroyum.** Fróðskapur. Faroe University Press. Tórshavn. 201 s.

Bogen giver et overblik over sønavne på Færøerne og er et bidrag til færøsk stednavneforskning med fokus på navnetraditioner, former og etymologi.

# Grønland

*Beatrine Heilmann*

## Rapporter:

Uddannelsesstyrelsen, 2021: **Folkeskolen i Grønland**. Publikationen bygger på kommunernes/skolernes statistiske indberetning for skoleåret 2020/21 til Uddannelsesstyrelsen. Publikationen kan hentes fra: <https://iserasuaat.gl/-/media/iserasuaat/folkeskole/7-publikationer/folkeskolen-i-groenland/kalaal-lit-nunaanni-atuarfik-2020-2021----folkeskolen-i-grnland-2020-2021.pdf>

Øgaard, Anders, 2021: **Kenyapakken i Sydgrønland**. Ilisimatusarfik, Nuuk, Grønland. Rapporten handler om fjernundervisning i engelsk på to bygdeskoler i Sydgrønland.

## Artikler:

Kristensen, Tikaajaat, 2021: **Sprog i fokus: Kalaallisut (grønlandsk)**. *Sprog I Norden*, 49(1), 110–116. Hentet fra <https://tidsskrift.dk/sin/article/view/114951>

Mathiassen, Ivalu, 2021: **Læsestrategier hos grønlandske elever**. IN: *Læsepædagogik*, nr. 1, vol.: 69. årgang, februar 2021, <https://www.laeseapaed.dk/laeseapaedagogik-nr-1-2021/>.

I denne artikel beskrives ph.d.-projektet “Elevers stavelses- og morfemiske læsestrategier på yngstetrinnet”. En undersøgelse af læseundervisningen i didaktiske perspektiv i faget grønlandsk, hvor eyetracking bliver brugt en del af empirisamlingen.

Nielsen, Flemming, 2021: **Hans Egede og den grønlandske bibel**. Udgivelsessted på nettet: [www.bibelselskabet.dk](http://www.bibelselskabet.dk). Publikationen er en kort oversigt over Hans Egedes første grønlandske tekster og hans første forsøg på at oversætte en længere tekst fra Bibelen, nemlig 1 Mosebog kap. 1-11. Der gives nogle eksempler på oversættelsesproblemerne og på, hvordan Hans Egede har påvirket det grønlandske sprog.

Nielsen, Flemming, 2021: **Vejen til himlen: den tidlige grønlandske katekismus**. IN: *Tro og samfund i 300-året for Hans Egedes ankomst til Grønland*. Editor: Aage Rydstrøm-Poulsen; Gitte Adler Reimer; Annemette Nyborg Lauritsen. Aarhus Universitetsforlag, Århus, Danmark. ISBN: 9788772194561.

Nielsen, Flemming, 2021: **Verbumtyper og diateser i grønlandsk – en skitse til syntaktisk karakteristik af grønlandske verber og en oversigt over det grønlandske diatesesystem.** IN: *Grønlandsk Kultur- og Samfundsforskning*, vol. 2020-21. Ilisimatusarfik og Det Grønlandske Selskab, Nuuk & København, Grønland & Danmark.

Trondhjem, Naja Blytman, 2021: **Greenlandic grammar.** IN: *Handbook of Inuit, Yupik and Unangan Languages.* Oxford University Press.

### **Bøger:**

Dalager, Maren Louise, 2021: **Nipit pinngorfii.** 1. udgave. ISBN: 9788772083025, 24 sider, Ilinniusiorfik Undervisningsmiddelforlag, Nuuk, Grønland.

Nielsen, Flemming A. J., 2021: **Vestgrønlandsk grammatik.** 2. udgave, 1. oplag, 363 sider. BoD & LearnGreenlandic. ISBN: 9788743027768. Bogen handler om grønlandsk sprog og grammatik.

Tobiassen, Johanne B., 2021: **Sprogpolitik i det politiske område: En undersøgelse om hvordan partierne udvikler formålsværdier også indenfor de sproglige aspekter.** 26 sider, Ilisimatusarfik, Nuuk, Grønland.

## Island

*Ágústa Þorbergsdóttir*

**Anna Vilborg Einarsdóttir, Ágústa Þorbergsdóttir og Sigríður Sigurðardóttir. Ráðandi tungumál í íslenskri ferðaþjónustu.** Háskólinn á Hólum. 53 s. Publiceret på: [https://www.holar.is/static/files/Ferdamaladeild/Utgefird/radandi\\_tungumal\\_i\\_isl\\_ferda.pdf](https://www.holar.is/static/files/Ferdamaladeild/Utgefird/radandi_tungumal_i_isl_ferda.pdf)

En rapport om det islandske sprogs status inden for turistindustrien i Island.

**Ágústa Þorbergsdóttir, Salka Guðmundsdóttir og Silja Bára Ómarsdóttir, 2021. Íðorðasafn í alþjóðastjórnámálum og stjórn málafræði.** Reykjavík: Stofnun Árna Magnússonar í íslenskum fræðum. Reykjavík. 56 s.

En islandsk-engelsk termsamling over statskundskab og international politik. Alle begreber er defineret.

**Ágústa Þorbergsdóttir og Steinþór Steinsgrímsson, 2021: Orð ársins 2020: Sóttkví.** I: Húgrás 20. januar 2021. Publiceret på: <https://hugas.is/2021/01/ord-arsins-2020-sottkvi/>.

En artikel om valget af årets ord 2020. Sóttkví ('karantæne') blev kåret som årets ord med henvisning til, at det har været særligt fremtrædende i årets løb.

**Ásgrímur Angantýsson, Sigríður Sigurjónsdóttir og Eiríkur Rögnvaldsson, 2021: Málið er: greinasafn 1980–2020.** Reykjavík: Háskólaútgáfan. 522 s.

*Málið er* er et festskrift til Höskuldur Þráinsson, professor i islandsk lingvistik ved Islands Universitet i anledning af hans 75-års fødselsdag. I bogen findes der et bredt udvalg af hans artikler fra en 40 års periode, 1980–2020. Festskriftet er redigeret af tre kolleger fra Islands universitet og de skriver også forord til bogen.

**Birna Arnbjörnsdóttir, 2021. Tölvutengd tungumálakennsla: Kennslufræði, árangur og möguleikar Icelandic Online námskerfisins.** Publiceret på: <http://millimala.hi.is/wp-content/uploads/2022/01/2021-2-Birna-Arn-bjo%CC%88rnsdo%CC%81ttir.pdf>

Artiklen handler bl.a. om mulighederne for at bruge Icelandic Online-kursmateriale ([www.icelandiconline.com](http://www.icelandiconline.com)) til at udvikle kurser på andre sprog.

**Guðrún Steinþórsdóttir og Sigrún Margrét Guðmundsdóttir (red.), Rit-  
ið. 3/20021. Tímarit Hugvísindastofnunar.** Reykjavík, Hugvísindastofnun  
Háskóla Íslands. 268 s. Publiceret på: [https://ritid.hi.is/index.php/ritid/issue/  
view/14](https://ritid.hi.is/index.php/ritid/issue/view/14)

Dette nummer af Ritið handler om indflydelsen af digital teknologi og medier på det islandske sprog og kultur og gæstredaktører for numret er Eiríkur Rögnvaldsson og Sigríður Sigurjónsdóttir. Numrets artikler er: Iris Edda Nowenstein og Sigríður Sigurjónsdóttir: Stafrænt málsambýli íslensku og ensku [Digital sprogkontakt mellem islandsk og engelsk], Hildur Hafsteinsdóttir, Sigríður Sigurjónsdóttir og Þorbjörg Þorvaldsdóttir: Íslenskur hreimur í málsambýli [Islandsk accent i forbindelse med sprogkontakt], Ragnheiður Jónsdóttir og Helga Hilmisdóttir: Leikjatölvur og orðaforði unglinga [Spillecomputer og teenageres ordforråd], Finnur Friðriksson og Ásgrímur Angantýsson: „Þá leyfir maður sér liggur við að hlaupa tungumálanna á milli“ [Om sprogveksling mellem islandsk og engelsk på samfundsmedier hos elever på gymnasial uddannelse], Agnes Sólmundsdóttir, Dagbjört Guðmundsdóttir, Lilja Björk Stefánsdóttir og Anton Karl Ingason: Vondar vélþýðingar [Dårlige maskinoversættelser] og Ari Páll Kristinsson: Rannsóknir á stafrænu málsambýli varpa nýju ljósi á meginþræði íslenskrar málstefnu. [Forskninger i digital sprogkontakt kaster nyt lys over hovedtrækkene i islandsk sprogpolitik.]

**Helga Hilmisdóttir (red.), 2021: Orð og tunga. Nr. 23.** [Ord og sprog. Tidsskrift udgivet af Árni Magnússon-instituttet for islandske studier.] 152 s.

Årgangens artikler er: Ari Páll Kristinsson: Innri breytileiki og málsnið [intra-speaker variation og sprogligt register], Ásta Svavarsdóttir: Málþróun og samfélagsbreytingar á síðari hluta 19. aldar. [Sprogudvikling og samfundsændringer], Þorsteinn G. Indriðason: Af hverju *góðlátlegur* en ekki *\*góðleglátur*? [Om suffikset *-látlegur*], Ágústa Þorbergsdóttir: Staða íslensku á sviði fjármála [Om islandsk sprogs status på finansområdet].

Bindet indeholder desuden to kortere bidrag: Svavar Sigmundsson skriver om kælenavne i en samling af genealogier og Jóhannes B. Sigtryggsson skriver om bindestreg.

**Höskuldur Þráinsson, 2021: Handbók um málfræði. Tredje udgave.** Reykjavík: Mál og menning. Reykjavík. 368 s. Redaktion: Árni Árnason, Heimir Pálsson og Laufey Leifsdóttir.

Dette er en opslagsbog om islandsk grammatik for lærere og elever. Bogen



er todelt. I den første del er en liste med kortfattede definitioner af grammatiske termer. Den sidste del har syv kapitler, der handler om udtale, fleksion, orddannelse, syntaks, semantik, sprogrøgt og sproghistorie.

**Íslenskt mál og almenn málfræði Nr. 43.** [Íslandsk sprog og generel lingvistik'. Tidsskrift udgivet af den islandske lingvistikforeningen] 2021. Red. Ásta Svavarsdóttir, Einar Freyr Sigurðsson og Þórhallur Eyþórsson. 208 s.

Dette nummer indeholder fire artikler, bl.a. om islandsk syntaks, sproghistorie og morfologi: Þorgeir Sigurðsson: Haustlǫng. How syntax and rhyme interact in early Old Norse *dróttkvætt*. Heimir F. Viðarson: Sögulegar breytingar á orðaröð með tveggja andlaga sögnum í íslensku. Þorsteinn G. Indriðason: Leitin að stofninum. Um stofnsamsetningar og samanburð við eignarfallssamsetningar. Veturliði Óskarsson og Þorsteinn G. Indriðason: Eingarfallssamsetningar með sterkum lýsingarorðum í fyrri lið. Söguleg og samtímaleg úttekt.

**Jóhannes Bjarni Sigtryggsson, 2021: Íslensk réttiritun.**

Jóhannes Bjarni har publiceret vejledninger om islandsk retskrivning, fremstillet som en webside tilpasset et online miljø: <https://rettritun.arnastofnun.is/>. På denne måde er informationerne mere tilgængelige.

**Katrín Axelsdóttir, Veturliði Óskarsson og Þorsteinn G. Indriðason, 2021. Möggubrár heklaðar Margréti Jónsdóttur sjötugri 21. mars 2021.**

Reykjavík: Rauðhetta útgáfufélag. 180 s. Publiceret på: [https://ritaskra.arnastofnun.is/media/skraning\\_pdf/moggubrar.pdf](https://ritaskra.arnastofnun.is/media/skraning_pdf/moggubrar.pdf)

*Möggubrár* er et festskrift til Margrét Jónsdóttir, professor i islandsk sprog. Det er redigeret af tre af Margréts kollegaer og udsendt i anledning af hendes 70-års fødselsdag 21. marts 2021.

Festskriftet er skrevet af venner og kollaeger og det indeholder 37 kortere bidrag og artikler om forskellige sproglige emner.

**Kristjana Vigdís Ingvadóttir, 2021. Þrautseigja og mikilvægi íslenskrar tungu. Um notkun dönsku og erlend áhrif á íslensku.** Reykjavík: Sögufélag. 320 s.

Bogen handler om hvilken effekt det havde på det islandske samfund at være under dansk herredømme (på det 18. og 19. århundrede). Kampen for det islandske sprog behandles og sprogbrugen på Island studeres på en systematisk måde ved at undersøge amtets embedsmænds korrespondance mellem embedsmænd.

**Málfregnir 29.** Islandsk Sprognævn's tidsskrift. 2021. 29 s.

I *Málfregnir 29* publiceres resolutionen om det islandske sprogs status år 2021 og også fem foredrag fra symposiet om sprogrøgt som blev afholdt 30. september 2021. Tidsskriftet er tilgængeligt på <http://islenskan.is/images/Malfregnir/malfregnir-2021-29-1-tbl.pdf>

**Orðlof veitt Þorbjörgu Helgadóttur sjötugri 18. maí 2021.** Reykjavík: Menningar- og minningarsjóður Mette Magnussen. 82 s.

*Orðlof* er et festskrift til Þorbjörg Helgadóttir, udgivet i anledning af hendes 70-års fødselsdag. Festskriftet indeholder 29 artikler, mange om filologi og leksikografi.

**Páll Björnsson. Ættarnöfn á Íslandi.** Reykjavík: Sögufélag. 324 s.

Historikeren Páll Björnsson diskuterer kontroversielle emner vedrørende efternavne i Island og knytter dem til historiske udviklinger, f.eks. til dannelsen af byområder og fremkomsten af nationalistiske bevægelser i tidligere tider.

**Ragnheiður Jónsdóttir. Enska í slangurorðaforða unglina á Íslandi.** Skírnir 195, s. 379–408.

Denne artikel af Ragnheiður Jónsdóttir handler om engelske slangord hos teenagere.

**Sálubót. Afmælisrit til heiðurs Jörgen Pind.** Reykjavík: Háskólaútgáfan. 248 s.

Festskrift til Jörgen Pind, professor i psykologi, i anledning af hans 70-års fødselsdag. Redigeret af Árni Kristjánsson, Heiða Margrét Sigurðardóttir og Kristján Árnason. Artiklene i festskriftet er skrevet af Jörgen Pind og hans kollegaer inden for psykologi og lingvistik. En række emne diskuterer, herunder dysleksi og visuel perception, psykologiens historie og oprindelse, klinisk psykologi, islandsk i udlandet, leksikografi og sprogrøgt.

## Norge

*Gunnbild Wiggen*

**Budal, Ingvil Brügger, Karine Stjernholm og Bernt Ø. Thorvaldsen.**

2021. *Norsk språkhistorie for lærarar*. Oslo: Fagbokforlaget.

Denne boka gjev ei grunnleggjande innføring i norsk språkhistorie og er skreddarsydd for grunnskulelæraren. Forfattarane fortel historia om norsk språk og skildrar dei viktigaste endringane, frå urnordisk, via norrønt og mellomnorsk og fram til moderne norsk. Språkleg mangfald, variasjon og endringar i fortid og samtid står sentralt i fagfornyinga. I boka er det lagt stor vekt på å visa skulerelevansen til fagstoffet og korleis dette kan bli formidla i klasserommet.

**Fridtun, Kristin.** 2021. *Kunsten å irritera seg over språkfeil*. Oslo: Samlaget.

Overraskande bok om velkjent irritasjonsmoment.

Me kommuniserer meir gjennom skrift enn nokon gong, samstundes som evna og viljen til å fylgja rettskrivingsreglane varierer.

Nokre menneske blir svært opprørte når dei støyter på skrivefeil, medan andre tek det heilt med ro. Kvifor reagerer me så ulikt? Kva hender i hjernen når me støyter på skrivefeil? Har personlegdomen vår noko å seia for korleis me reagerer? Og korleis går me fram dersom me vil bli kvitt irritasjonen?

Språkvitar Kristin Fridtun har sett seg føre å svara på desse spørsmåla. Hen grev i faglitteraturen, dukkar ned i kommentarfelta, fretter ut irritable språkbrukarar og granskar irritasjonen frå ulike sider. Hen finn ut at personlegdom har ein del å seia, men at reaksjonsmønsteret vårt likevel ikkje er hogge i stein. Dei aller fleste av oss kan få bukt med skrivefeilirritasjonen, berre me går inn for det.

Kristin Fridtun (f. 1987) frå Elverum er utdanna filolog, bur på Lillehammer og er forfattar på heiltid. Hen har skrive fleire bøker om språk, som *Norsk etymologisk oppkok* (2012), *Nynorsk for dumskallar* (2013) og *Språket er ei bukse* (2019). Fridtun vann Bokhandelens sakprosapris i 2016 og Språkprisen i 2019.

**Faarlund, Jan Terje, Svein Lie og Kjell Ivar Vannebo.** 2021 [ny utgave].

*Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

*Norsk referansegrammatikk* fra 1997 er en standardgrammatikk over moderne norsk språk – bokmål og nynorsk. Det er norgeshistoriens største og

mest omfattende grammatikk. Boka gir en sammenhengende beskrivelse av norsk språk, men kan også brukes som oppslagsverk. Framstillingen er basert på en samling og organisering av store kunnskapsmengder og på ny forskning i nordisk og allmenn språkvitenskap. Grammatikken vil derfor være et nyttig hjelpemiddel for alle som har bruk for kunnskap om dagens norske språk.

**Haugen, Tor Arne, Svenn-Arve Myklebost, Stig Jarle Helset og Endre Brunstad.** 2021. *Språk, tekst og medvit.* Oslo: Cappelen Damm Akademisk.

Dette nummeret av skriftserien Skriftkultur viser korleis det tverrfaglege skriftkulturelle feltet kan opne nye perspektiv på grunnleggande spørsmål. Eit avgjerande grunnlag for all forskning og vitenskap er at han er skriftbasert. Det gjeld ikkje berre i humaniora, men i alle vitenskaplege disiplinar.

Vitenskaplege teoriar og tenkjemåtar er språklege konstruksjonar i kulturelle fellesskap, og det er i skriftleg form vitenskapleggjinga skjer. Grunnlagsproblem i utforsking av språket og skriftinga er dermed samstundes grunnlagsproblem for andre skriftspråklege praksisar, anten dei er vitenskaplege eller meir generelle. Tilhøva mellom språket, tekstane det gjev grunnlag for, og medvitet til språkbrukarane handlar nettopp om slike grunnproblem. Kva er grunnlaget vårt for å forstå, beskrive og forklare desse tilhøva, og på kva måtar gjev ulike fag og disiplinlar ulike inngangar til og svar på slike spørsmål?

I dette nummeret undersøker sju forskarar tilhøvet mellom språk, tekstar og medvit med utgangspunkt i ulike fag innan human- og samfunnsvitenskapane: språkvitenskap, fagdidaktikk, psykolingvistikk, språkfilosofi, estetisk teori, tekst- og litteraturvitenskap og religion. Publikasjonen rettar seg mot studentar, forskarar og andre som er opptekne av dette viktige feltet.

**Hem, Erlend og Magne Nylenna** (red.). 2021. *Helsespråk. Michael 2021*; 18: Supplement. Oslo: Det norske medicinske Selskab.

**Hårstad, Stian.** 2021. *Nabospråk og nabospråkundervisning.* Oslo: Cappelen Damm Akademisk.

Er det likegyldig om Pippi snakker svensk eller norsk på tv?

Er det greit at du må snakke engelsk når du bestiller pølse i Legoland?

Hva kan gjøres for at de skandinaviske språkene fortsatt skal være gangbare i hele Skandinavia?

Hovedsiktet med denne boka er å tilby en innføring i temaet nabospråk med vekt på undervisning i emnet. Arbeid med svensk og dansk er en del av læreplanens mål, men dessverre ser det ut til å bli nedprioritert av en del norsklærere. Denne boka viser hvorfor arbeid med nabospråkforståelse er

viktig og nyttig, hva de språklige utfordringene består i, og ikke minst hvordan man kan arbeide med nabospråk på varierte og læringsfremmende måter.

Nabospråk og nabospråkundervisning er rettet mot lærere og lærerstudenter som har – eller kommer til å få – grunnskolen som sin arbeidsplass.

**Hårstad, Stian, Brit Mæhlum og Rikke van Ommeren.** 2021. *Blikk for språk. Sosiokulturelle perspektiver på norsk språkvirkelighet*. Oslo: Cappelen Damm Akademisk.

Hvorfor slutter noen å bruke dialekten sin selv om Norge er kjent for å være et «dialektparadis»?

Hvorfor blir «knoting» ofte oppfattet som negativt?

Hvorfor vinner bokmål terreng i forhold til nynorsk når skriftformene er offisielt likestilte?

Hvorfor blir noen stavemåter sett ned på selv om de har vært tillatt i mer enn hundre år?

Denne boka gir innsikt i en essensiell dimensjon ved språk, nemlig at språklige valg uttrykker sosiale og kulturelle posisjoner.

Med blikk for språkets ideologiske aspekter og sosiale funksjoner ser forfatterne på fenomener som standardisering, dialektnivellering, «knoting» og multi-etnolekt, språkbruk i digitale tekster og reklame, og forholdet mellom norsk og andre språk fra norsk språkhistorie, som samisk, dansk, tysk og engelsk.

*Blikk for språk. Sosiokulturelle perspektiver på norsk språkvirkelighet* er skrevet for studenter i norsk- og nordiskfaget og passer på bachelor- og masternivå. Leserne blir kjent med sentrale begreper fra sosiokulturell lingvistikk og får innsikt i noen av språkvitenskapens grunnlagsspørsmål. Framstillingen er særlig aktuell for lærerstudenter som skal bidra til elevers danning i et demokratisk og flerkulturelt samfunn.

**Jahr, Ernst Håkon og Gro-Renée Rambø** (red.). 2021. *Norsk språk i rom og tid*. Oslo: Novus.

Forskningsgruppa i språkhistorie (HistLing) ved Fakultetet for humaniora og pedagogikk, Universitetet i Agder, legg med dette fram skrift nr. 1 frå miljøet.

Ein tanke bak denne samlinga av artiklar er å vise fram noko av mangfaldet og spennvidda i ei av dei største forskingsgruppene ved fakultetet. Etter ei omfattande internasjonal evaluering av humanistisk forskning (HumEval), som Forskningsrådet gjorde ved norske universitet og høøgskolar med rapport i 2017 (Evaluering av humanistisk forskning i Norge), blei forskingsgruppa tildelt NFR-middel til vidare utvikling. Eit resultat av dette er no m.a. denne antologien.

Heilt bak i boka finn ein ei liste over medlemmene i gruppa med opplysningar om kva for område av språkvitskapen dei primært arbeider på. Gruppa samlar i alt 15 forskarar frå ulike område og særskilt innan diakron språkforskning. Dei fleste er knytte til Instituttet for nordisk og mediefag, to kjem frå Instituttet for framandspråk og omsetting. Gruppa har eit fleirtal kvinnelege forskarar, og det er òg med fleire aktive emeriti. 11 av forskarane i gruppa er representerte i denne antologien som forfattarar eller redaktørar.

**Jølle, Lennart, Ann Sylvi Larsen, Hildegunn Otnes og Leiv Inge Aa.** 2021. *Morsmålsfaget som fag og forskningsfelt i Norden*. Oslo: Universitetsforlaget.

Morsmålsfaget er stort og mangfoldig og omfatter mange deldisipliner. Det utgjer et «fagmassiv», som lærere, lærerutdannere og forskere må orientere seg i. I denne antologien er det samlet 13 artikler som demonstrerer eksemplar på hvordan det både kan undervises i og forskes på morsmålsfaget. Boka inneholder artikler innanfor språkdidaktikk, litteraturredidaktikk og skrivedidaktikk og har empiri fra alle nivå i utdanningssystemet i Norge, Sverige, Danmark og Finland. Antologien speiler også at morsmålsfaget er i stadig bevegelse, og nye disipliner kommer til. Derfor har vi for eksempel med bidrag som digital skriving i sosiolingvistisk perspektiv, sanglyrikk som multimodalt uttrykk og pedagogisk transspråking. Antologien har blitt til i etterkant av en nordisk konferanse om morsmålsdidaktikk ved NTNU, Trondheim, i 2019. Konferansekomiteen var den samme som redaksjonen for denne antologien. Både konferansen og boka inngår i en rekke slike arrangement som går på omgang i morsmålsdidaktikkmiljøene i Norge, Sverige, Danmark og Finland. Det felles nordiske nettverket har navnet Nordisk nettverk for morsmålsdidaktisk forskning (NNMF).

**Karlsen, Helene Urdland.** 2021. *Norsk ordbok – nynorsk*. Oslo: Cappelen Damm Undervisning.

*Norsk ordbok – nynorsk* inneheld det sentrale ordtilfanget i moderne nynorsk. Ordboka er laga spesielt for elevar i grunnskulen og i den vidaregåande skulen.

Cappelen Damms *Norsk ordbok – nynorsk* er den nyaste og mest oppdaterte ordboka i sitt slag. Ordboka tar utgangspunkt i moderne norsk og er grundig gjennomarbeidd med tanke på lesbarheit, eit aktuelt ordtilfang og brukaranes behov for ei påliteleg og praktisk ordbok.

**Kjelaas, Irmelin og Kristin Fagerheim.** 2021. *Integrert språk- og fagopplæring for nyankomne andrespråkselever*. Oslo: Universitetsforlaget.

Nyankomne andrespråkselever står overfor en stor læringsoppgave: De skal

tilegne seg det norske språket samtidig som de skal fortsette sin faglige utvikling – på norsk. Dette krever at de får ei tilpassa opplæring både i språk og fag, at språk- og fagopplæringa ses i sammenheng, og at opplæringa er ressursorientert og tar utgangspunkt i elevenes erfaringer, kunnskaper og flerspråklige kompetanse. Denne boka gir ei teoretisk og didaktisk innføring i hvordan ei slik helhetlig og integrert opplæring kan planlegges, organiseres og gjennomføres. Forfatterne presenterer det sosiokulturelle teorigrunnlaget for slik opplæring, før de tar for seg det didaktiske arbeidet med ordforråd, muntlighet, lesing og skriving. Opplæring for nyankomne minoritetsspråklige elever er i fokus i boka. Den er også nyttig lesning for alle lærerstudenter og lærere som kommer til å få, eller har andrespråkselever i sitt klasserom.

**Kverndokken, Kåre, Jannike Ohrem Bakke og Ingvil Brügger Budal** (red.). 2021. *101 grep om grammatikk. Om språket som system og språket i bruk*. Oslo: Fagbokforlaget.

Hva er grammatikkens rolle i fagfornyelsen (LK20)? Hva er formålet med grammatikkundervisning, og hvordan lykkes man med den?

*101 grep om grammatikk* er den åttende boka i *101-serien*, og i likhet med de andre titlene legges det også her stor vekt på å forene teori og praksis. Boka har et knippe fagfelleverderte, vitenskapelige artikler som tematiserer grunnleggende spørsmål om grammatikk og grammatikkundervisning. Med denne boka får undervisere på ulike nivåer en didaktisk verktøykasse til egen grammatikkundervisning.

*101 grep om grammatikk* bidrar til ny forskning på feltet. Boka er en skattkiste for alle som skal jobbe med å utvikle elevers grammatiske kompetanse i hele det 13-årige skoleløpet. Målgruppen er lærere, studenter, lærerutdannere og forskere.

**Matre, Synnøve, Randi Solheim, Hildegunn Otnes, Kjell Lars Berge, Lars Sigfred Evensen og Ragnar Thygeson.** 2021. *Nye grep om skriveopplæringa*. Oslo: Universitetsforlaget.

Korleis kan ei felles forståing av skriving og tydelege forventningar om skrivekompetanse gi betre skriveopplæring? Og korleis kan ein utnytte læringspotensialet som ligg i skriving i ulike fag og samanhengar? I fleire år har forskarar følgt ei stor gruppe lærarar og elevar i arbeidet deira med skriving og vurdering. Funna er interessante og oppsiktsvekkande. Denne boka presenterer bakgrunnen for det landsomfattande Normprosjektet, gjennomføringa, funn og erfaringar. Boka inneheld også praktiske ressursar for skriveopplæring og tekstvurdering og presenterer erfaringar med å ta desse i bruk. Lærar- og

elevstemmer kling saman med forskarstemmene gjennom heile boka, mellom anna i form av skriveoppdrag, elevtekstar frå ulike fag og samtalar om tekstar. Slik vender boka seg til ei brei lesargruppe: skriveforskarar, lærarutdannarar, skuleleiarar, utviklingsorienterte lærarar og ikkje minst lærarstudentar. Forskarane som har designa og gjennomført prosjektet, har i fellesskap forfatta boka, og redaktørane har skrive det heile saman. I eit etterord set Judy Parr frå New Zealand prosjektet inn i ein internasjonal kontekst.

**Monsen, Marte og Verónica Pájaro** (red.). 2021. *Andrespråkslæring hos voksne. Vitenskapelige innsikter og didaktiske refleksjoner*. Oslo: Cappelen Damm Akademisk.

Hvordan lærer voksne norsk som andrespråk? Hvordan bør språkopplæringen være?

De fleste utgivelsene innen norsk som andrespråk handler om barns læring av norsk i skolen og i barnehagen. Voksne innlærere har ikke fått like stor oppmerksomhet i forskningen. Voksne lærer imidlertid ikke språk helt på samme måte som barn, og voksnes språklæring skjer i andre kontekster og under andre betingelser. Derfor er det viktig med en oppdatert bok om andrespråkslæring hos voksne spesielt, både for forskningens del og for didaktisk refleksjon.

Denne boka presenterer ny kunnskap om andrespråkslæring og andrespråksdidaktikk for voksne. Sentrale temaer er morsmålsbegrepet, språklige repertoar, investering, identitet, ordforråd, syntaks, numerasitet, lese- og skriveutvikling, Norskprøven, språkkrav og jobbintervjuer.

Forfatterne er forskere innen flerspråklighet og språklæring, og de undersøker temaene fra ulike perspektiver; kritiske, sosiolingvistiske, sosiokulturelle og kognitive. De drøfter også hvilke konsekvenser de nye innsiktene bør få for norskopplæringen.

Boka er aktuell for alle som er involvert i andrespråksfeltet og språkopplæringen i norsk, som politikere, forskere, studenter og skolefolk. Den er særlig nyttig for lærere som underviser voksne i norsk som andrespråk.

**Maal og minne**. Vol 113, nr. 1 og 2 (2021; utgitt hhv. 8.6.2021 og 17.12.2021). Oslo: Novus.

*Maal og Minne* er Bymålslagets tidsskrift. Tidsskriftet ble grunnlagt i 1909 av Magnus Olsen og har til formål å bringe bidrag til belysning av bl.a. norsk språk og norske språkminner av enhver art, middelalderlitteratur, norske stedsnavn og folkeminner. Bidrag til tidsskriftet blir vurdert av redaktørene og av uavhengige konsulenter (*referees*). *Maal og Minne* kommer ut med to hefter



i året og trykker artikler og anmeldelser på norsk, svensk, dansk, engelsk og tysk.

**Norsk Lingvistisk Tidsskrift.** Vol 39, nr. 1 og 2 (2021; utgitt hhv. 2.7.2021 og 14.1.2022). Oslo: Novus.

*Norsk Lingvistisk Tidsskrift* (NLT) er et forum for norsk språkvitenskap og bringer lingvistiske artikler og bidrag på norsk innenfor alle grener av og områder i språkvitenskapen, også felt som faghistorie og lingvistikkrelatert vitenskapsteori. NLT er et fagfelleverdert tidsskrift og utgis med støtte fra Norges forskningsråd.

**Nygård, Mari.** 2021. *Grammatikk og grammatikkdidaktikk*. Oslo: Samlaget.

Grammatikarar er språkmekanikarar, og arbeid med grammatikk handlar om å forsøke å finne ut korleis språkmaskineriet fungerer, kva komponentar det har, og korleis dei verkar saman.

Barn tileignar seg grammatikken i morsmålet sitt heilt knirkefritt. Vi vil spørje korleis ein kan vekkje interesse for innsikt i språkstruktur, og korleis ein kan leggje opp til undring og utforsking av sitt eige morsmål, slik at grammatikkundervisninga blir spennande. Kunnskap om norsk språkstruktur og omgrepsapparat omkring dette er grunnleggjande for mange område innanfor norskfaget, mellom anna lese- og skriveopplæring, norsk som andrespråk, talemål og nynorskdidaktikk.

Denne boka inneheld både grammatikk og grammatikkdidaktikk. Det mest sentrale vil vere å gi ei grunnleggjande innføring i faget grammatikk. Her vil vi leggje vekt på innlæring av grammatisk terminologi og analysemåtar. I tillegg vil vi gjennom heile boka diskutere korleis ein kan nærme seg dette emnet saman med elevar, og også kvifor dette er eit viktig emne, både i seg sjølv og som byggestein for andre emne i norskfaget.

**Seljeseth, Ida.** 2021. *Klart språk og retorisk medborgerskap*. Doktorgradsavhandling. Oslo: Universitetet i Oslo.

Klart språk, eller plain language, er en internasjonal innsats for å gjøre institusjonelle tekster, som for eksempel vedtaksbrev eller helseinformasjon, mer forståelige. En vanlig begrunnelse for å jobbe med klart språk er at uklart språk er et demokratisk problem, og til hinder for rettssikkerheten.

Tanken om at uklart språk hemmer demokratiet og rettssikkerheten, har resultert i en oppfatning om at motsetninga, klart språk, fremmer det. I avhandlinga argumenterer Seljeseth for at dette ikke er en selvfølge. Klart språk er en forutsetning for deltakelse i samfunnet, men det er også en forutsetning

for språklig maktutøvelse. Forskjellen ligger i *hva* som gjøres klart. Hvis klar-språkinnstasen i hovedsak er retta mot å gi tydelige beskjeder, kan klart språk like gjerne være hemmende for demokratiet og rettssikkerheten.

Det betyr at klarspråkprosjekter som har som formål å fremme demokratiet og rettssikkerheten må ha et bevisst forhold til makt. I noen tekster kan økt forståelighet gjøre at flere lesere tar informerte valg. I andre tekster kan det hende at arbeidet med å øke forståeligheten går på bekostning av nyansene på en sånn måte at flere gjør det de får beskjed om, uten å ha noe bevisst forhold til at de kunne gjort noe annet.

For å sikre at klarspråkprosjektene faktisk fremmer, og ikke hemmer, demokratiet og rettssikkerheten, foreslår Seljeseth at *retorisk medborgerskap* etableres som et selvstendig mål for klarspråkarbeid. Retorisk medborgerskap er knytta til medborgernes retoriske handlekraft – i offentligheten og i møte med det offentlige. Det er knytta til evnen til å ytre seg relevant, som igjen er knytta til forståelse. I tillegg påvirkes retorisk handlekraft av muligheten til å bli hørt – i denne sammenhengen offentlig sektors holdning til medborgernes ytringer.

**Språknytt.** Språkrådet. 2021.

<https://www.sprakradet.no/Vi-og-vart/Publikasjoner/Spraaknytt/>

**Statsspråk.** Språkrådet. 2021.

<https://www.sprakradet.no/Vi-og-vart/Publikasjoner/Statssprak/>

**Termposten.** Språkrådet. 2021.

<https://www.sprakradet.no/Vi-og-vart/Publikasjoner/termposten/>

## Sverige

*Jobanna Ledin*

**Aare, Cecilia, 2021: Reportaget som berättelse: en narratologisk undersökning av reportagegenren.** Doktorsavhandling, Stockholm: Stockholms universitet. 301 s.

Avhandlingens syfte är att kartlägga reportagegenren utifrån en narratologisk modell som författaren utvecklat. Med utgångspunkt i modellen synliggör Aare hur en berättare, upplevande karaktärer och regisserande instans fyller olika funktioner som, även de på olika vis, påverkar vad reportaget förmedlar till läsaren. De frågor avhandlingen besvarar är på vilka sätt ett reportage kan berättas och hur narrativiteten mellan olika undergrupper varierar, samt hur ett narrativt engagemang som speglar olika yrkesrollers engagemang skapas.

**Aktürk Drake, Memet m.fl. (red.), 2021: Språk och stil: tidskrift för svensk språkforskning. Tema Flerspråkighet.** Uppsala: Institutionen för nordiska språk vid Uppsala universitet. Publicerad på <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-434146>. 231 s.

Den årliga tidskriften Språk och stil publicerar språkvetenskapliga artiklar och recensioner av språklitteratur. I detta temanummer av tidskriften presenteras aktuell forskning om flerspråkighet av forskare verksamma i Sverige. De sju bidragen i temanumret behandlar tillsammans flerspråkighetsforskningens historik i Sverige, finskans roll och status i Sverige under 2000-talet, flerspråkig utveckling hos barn, avancerad andraspråksinlärning hos vuxna, relationen mellan flerspråkighet och tanke, språkliga praktiker och språkattityder med koppling till ungdomar i förorten, samt relationen mellan sexualitet och språk i nutida migrationskontexter.

**Andersson, Lars-Gunnar, 2021: Göteborgslexikon.** Första upplagan, Stockholm: Morfem. 255 s.

Göteborgslexikon är en ordbok med enbart göteborgska ord och uttryck. Boken omfattar runt tusen ord och uttryck med betydelser, användning, uttal och ursprung. Lexikonet har liknats vid ett ordalbum för infödda och en inbjudan till ett nytt språkligt landskap för inflyttade. Vissa ord är allmänt bekanta, som a-män, bamba, dänka, exter, käckt, la, läska, måckli, svale med flera. Andra är inte lika välkända, som flane, gigga, hånka, joggens och tjingsa. Ytterligare andra ord är numera sällsynta, som apparell, golja, klissa, tjala och ånk.

Slutligen är en del ord så självklara att man inte tänker på dem som dialektord, som att man säger tissa för tisdag, jämte för bredvid och mårffar för morfar.

**Ask, Sofia, 2021: Skrivande polis.** Andra upplagan, Lund: Studentlitteratur. 118 s.

För de flesta poliser har skrivandet en central roll. Texter ingår i både förundersökningar och längre utredningar för olika läsare. Myndighetsutövningen som polis ställer vidare krav på ett vårdat och tydligt språkbruk. I boken behandlas flera infallsvinklar på skrivande som kan kopplas till yrkesrollen polis, som att göra tal till text, att förhålla sig objektiv och att skriva för olika mottagare. Författaren resonerar om hur formuleringar och ordval styr läsaren samt visar hur texter bör utformas för att vara funktionella i utredningsprocessen. Här finns också handfasta råd om vad som är språkligt viktigt att tänka på i texter som anmälan, förhör och PM. Boken riktar sig till både polisstudenter och yrkesverksamma poliser.

**Ask, Sofia & Byrman, Gunilla (red.), 2021: HumaNetten Nr 46, tema: Språk och rätt i teori och praktik.** Växjö: Fakulteten för konst och humaniora, Linnéuniversitetet. Publicerad på <https://open.lnu.se/index.php/hn/issue/view/178>. 119 s.

Vårnumret av HumaNetten utgår från en konferens på temat "Språk och rätt i teori och praktik" som hölls vid Linnéuniversitetet 2019. De frågor som artiklarna tar upp bekräftar den viktiga roll humaniora har i olika samhällsfrågor, inte minst knutet till detta tema. Stefan Holgersson ger i sitt bidrag en inblick i hur Polismyndigheten kommunicerar. Ulla Tiilikäs visar hur den offentliga makten, förvaltningen och servicen grundar sig på språkanvändning. Laura Ervo, Lukas Kindberg, Marcel Danesi och Kerstin Nordlöf diskuterar metoo-rörelsen som en sociopolitisk form av e-diskurs i termer av effektivitet och diskursiva former. Corina Löwe och Beate Schirmacher skriver om hur journalister rapporterar från rättegångar. Aud Anna Senjes artikel berör bl.a. klarspråksarbete i norska lagar, och Cecilia Wadensjö, Hanna Sofia Rehnberg och Zoe Nikolaidou följer en asylsökande i mötet med text i en tolkad asy-lintervju.

**Blomström, Vendela & Persson, Catarina, 2021: Muntlig interaktion i akademiska sammanhang.** Andra upplagan, Lund: Studentlitteratur. 268 s.

Som student ställs man inför många språkliga utmaningar – det gäller inte minst alla studenter med svenska som andraspråk. Boken syftar till att stötta dem i muntliga sammanhang, och innehåller både beskrivande textavsnitt

och praktiska övningar. Den passar för kurser på universitet och högskola liksom på kurser i svenska som andraspråk på gymnasium och inom vuxenutbildning.

**Blomström, Vendela & Wennerberg, Jeanna, 2021: Akademiskt läsande och skrivande.** Andra upplagan, Lund: Studentlitteratur. 356 s.

Studier på universitet och högskola är en utmaning för många. Många studenter vet inte hur man ska tackla de högt ställda förväntningarna på att planera och studera effektivt. Andra saknar lässtrategier för att förstå vetenskaplig text, liksom kunskap om hur man anpassar sitt skrivande till den akademiska genren och stiltnivån. Boken innehåller resonerande textavsnitt varvat med praktiska övningar och konkreta råd. Författarna förklarar hur man utvecklar sitt skrivande och sin språkfärdighet genom respons från andra, samt lyfter fram utmärkande drag i det akademiska språkbruket.

**Brandtler, Johan & Kalm, Mikael, 2021: Nyanser av grammatik: gränser, mångfald, fördjupning.** Första upplagan, Lund: Studentlitteratur. 338 s.

Grammatisk fördjupning innebär till en del att analysera de "sanningar" som den grundläggande grammatiska beskrivningen förmedlar. I den här boken problematiserar olika författare i arton kapitel en rad föreställningar och aspekter av svenskans grammatik, liksom grammatikbeskrivningens praktiska och teoretiska nytta: Hur många ordklasser finns det – egentligen? Hur ska man skilja på adverb och adjektiv, på huvudverb och hjälpverb? Var går gränsen mellan lexikon och grammatik, och hur ser samspelet ut? Boken riktar sig främst till studenter på fortsättningskurser i svenska och nordiska språk, inklusive blivande ämneslärare.

**Brändén, Henrik, 2021: Populärvetenskapens poetik och retorik: om konsten att förklara.** Lund: Celanders förlag. 139 s.

Genom åren har vetenskapsskribenten Henrik Brändén i egenskap av författare och föreläsare haft målet att skapa intresse och förståelse för komplicerade biologiska sammanhang. Den här boken handlar däremot om texter, om allt från hur man bygger upp en text rent tekniskt till hur man lär sig se texters mer subtila och konstnärliga sidor. Boken riktar sig till den som vill bli bättre på att skriva populärvetenskapligt, men också till den som vill veta mer om skönlitterärt skrivande.

**Dahl, Alva, 2021: Slå rot i förvandlingen: Gunnel Vallquists liv och livshållning.** Lund: Ellerströms. 389 s.

Författaren, översättaren och akademiledamoten Gunnel Vallquist (1918–2016) var länge ett namn i offentligheten, dels som översättare av Marcel Prousts, dels som essäist och debattör i kyrkliga frågor. Hon satte upp stränga principer för bildning och livsföring, sökte sina andliga rötter i Frankrike, Italien och Ryssland, och förstod att det mest avgörande var att lära sig bejaka tillvarons paradoxer: "Att gå in i motsägelsen, att vila i spänningpunkten, att slå rot i förvandlingen", som hon skrev i betraktelsesamlingen *Steg på vägen*. I denna biografi följer Alva Dahl de litterära, teologiska och livspraktiska trådarna i en komplex människas liv, samtidigt som hon breddar bilden av det svenska 1900-talets kulturliv.

**Devine, Josefin, 2021: Bygden, byarna och buan: studier av bebyggelsenamnen i Hackås och Ovikens socknar, med ett särskilt avsnitt om fäbodnamnen.** Doktorsavhandling, Umeå: Umeå universitet. 199 s.

Avhandlingen behandlar bebyggelsenamn från Hackås och Ovikens socknar i Jämtland. Den innehåller en analys av 79 enskilda namnartiklar med en fördjupning av 151 fäbodnamn. Syftet i studien har varit att besvara frågor om hur namnen är konstruerade, till vilket språkmaterial de är bildade, och vilken saklig grund namnen har. Studien ingår i serien *Sveriges ortnamn*, som ges ut av Institutet för språk och folkminnen.

**Douglah, Jessica, 2021: Dans, mobilitet, multimodalitet: en interaktionell studie av lärandesituationer i showdansklasser.** Doktorsavhandling, Stockholm: Stockholms universitet. 97 s.

Dansare använder sig av sin kropp för att förmedla en berättelse, en känsla eller ett budskap. Denna resurs kan även användas av danslärare i olika undervisningssituationer. Syftet med avhandlingen är att undersöka interaktion i en lärandesituation där många uttryckssätt används för att skapa en gemensam uppfattning om hur dansen ska framföras. Studien har dessutom ett metodologiskt syfte att bidra till utveckling av multimodal transkription av mobila aktiviteter.

**Flyman Mattsson, Anna, 2021: Andraspråkets grammatik.** Första upplagan, Lund: Studentlitteratur. 224 s.

Denna grundbok med tillhörande övningsbok belyser andraspråkets grammatik. Boken riktar sig till studerande på högskola och universitet som läser ämnet svenska som andraspråk. Den ger även stöd till lärare i svenska som andraspråk i grundskola och gymnasium.

**Flyman Mattsson, Anna & Håkansson, Gisela, 2021: Bedömning av svenska som andraspråk: en analysmodell baserad på grammatiska utvecklingsstadier.** Andra upplagan, Lund: Studentlitteratur. 160 s.

Många elever med ett annat modersmål än svenska har svårt att nå målen i skolans olika ämnen. Ofta beror det på att deras svenska inte utvecklats i samma takt som innehållet blir mer komplext i olika ämnen. I den här boken beskriver författarna grammatiska utvecklingsstadier i svenska och resonerar om hur man skapar bra förutsättningar för språkutveckling. Boken vänder sig till lärarstudenter och verksamma lärare med behov av ett analysverktyg för andraspråksutveckling.

**Forsberg, Camilla, 2021: Skrivandets gränser: Normering genom skrivdiskurser i tidig skrivundervisning.** Doktorsavhandling, Växjö: Linnéuniversitetet. Tillgänglig på Internet: <https://lnu.diva-portal.org/smash/get/diva2:1553153/FULLTEXT04.pdf>. 240 s.

Avhandlingen handlar om hur skrivundervisning ser ut i förskola, förskoleklass och grundskola upp till årskurs 3. Den bygger på observationer av skrivaktiviteter, dels för tre- till femåringar på en förskola, dels i en grundskola från förskoleklass upp till årskurs 3. I observationerna har primärt det som läraren lyfter fram och poängterar som viktigt noterats. Resultaten visar att eleverna driver den så kallade färdighetsdiskursen medan lärarna försöker stimulera barnens kreativitet. När eleverna efterfrågar skriv- och stavningsregler, är lärarnas intention att främja skrivandet i sig.

**Fridell, Staffan, 2021: Ordskloak: svenska ords släktskap och ursprung.** Stockholm: Kaunitz-Olsson. 359 s.

Detta är en etymologisk ordbok som visar vilka svenska ord som är besläktade med varandra och hur. Boken riktar sig till ordvitsare, korsordslösare och språkpoliser, personer som följer frågeprogram på teve, läser språkspalter i tidningar och lyssnar på programmet *Språket* i radio, kort sagt alla som gillar ord och vill veta mer. Författaren är professor emeritus i nordiska språk vid Uppsala universitet. Han är även etymologisakkunnig åt Svenska Akademiens ordbok.

**Grahn, Inga-Lill & Lindholm, Camilla (red.), 2021: Tala om kroppen: språkliga perspektiv på hälsa och sjukdom i den digitala eran.** Första upplagan, Stockholm: Morfem. 161 s.

Den här boken har kroppen i fokus: hur vi talar om den, upplever den och beskriver den. Med språkvetenskapliga metoder tar författare upp olika

perspektiv på språk och kropp, som hur vi söker kunskap om sjukdomar, hur vi använder metaforer för att tala om cancer, hur ett flerspråkigt samtal på mödrahälsovården kan se ut, och mer. Boken tar även upp den aktuella frågan vad som händer med kroppen i ett samtal där kroppen är frånvarande, som i ett digitalt möte. I boken medverkar språkforskare från Sverige och Finland.

**Hasan, Baki (red.), 2020: Språkrådets svensk-romska socialordlista.** Stockholm: Institutet för språk och folkminnen. Publicerad på <https://www.isof.se/stod-och-sprakrad/spraktjanster/svensk-romska-ordlistor>.

Språkrådets svensk-romska socialordlista innehåller uppslagsord och fraser på svenska och deras motsvarigheter på romska. Den nya utgåvan omfattar dialekterna kelderash, arli och lovani. Målgruppen för ordlistan är dels tolkar och översättare, dels personal inom social omsorg som kan lite romska eller ingen romska alls, men som ibland behöver använda romska i sitt arbete.

**Henderson, Fridha & Frid, Lina, 2021: Kreativt skrivande, Övningar – fakta – inspiration.** Didacta Förlag. 68 s.

Detta är en inspirationsbok i kreativt skrivande. Den innehåller ett femtiotal skrivövningar varvat med fakta om skrivprocessen och skrivtips kopplade till olika texttyper och framställningsformer: argumenterande text, berättande text m.fl. Boken passar den som går en skrivarkurs, läser svenska på gymnasienivå eller har skrivande som hobby.

**Holm, Anette, 2021: Glasklart! Skriv bättre på jobbet.** An1 Förlag. 85 s.

Det här är en bok till alla som skriver i jobbet och vill bli bättre på det. Boken ger råd om hur man lär sig att hitta rätt ordval, ton och stil. Det är ingen grammatikbok, utan handlar om skrivprocessen varvat med skrivtips. Boken tar upp exempel från olika genrer och texttyper, som säljande text, mejl, minnesanteckningar, beslutsunderlag, nyhetsbrev och sociala medier.

**Håkanson, Nils, 2021: Dolda gudar: en bok om allt som inte går förlo-  
rat i en översättning.** Stockholm: Nirstedt/litteratur. 356 s.

Med kunskap och många exempel från forntid till nutid visar författaren hur synen på översättning har varierat mellan epoker. Boken behandlar allt från reformationens betydelse för översättningen av Bibeln till svenska, till hur politik och ideologi på olika sätt villkorar översättningsverksamhet. Den tar även upp frågan om det finns ett korrekt sätt att översätta. Boken utsågs till Årets fackbok på Augustpriset 2021.



**Häger, Björn, 2021. Intervjuteknik: en handbok.** Tredje upplagan Stockholm: Morfem. 276 s.

Boken går igenom grunderna i intervjuteknik, belysta med konkreta exempel. Man får veta hur man får folk att öppna sig, hur man själv blir en bättre lyssnare och hur man lär sig hantera svårintervjuade personer. Boken vänder sig till såväl studenter som erfarna journalister. Den kan läsas av den som använder intervjun som redskap i sitt jobb, men även av den som bara vill veta mer om vilka villkor som styr ett samtal.

**Johansson, Sverker, 2021: På vandring i språkens fotspår.** [Elektronisk resurs]. Natur & Kultur. 430 s.

Det här är boken för dig som undrar över varför det finns så många språk i världen, och vilka mönster som finns bakom myllret. Den tar upp frågor som varför det talas tusen språk på Nya Guinea, men bara ett på Madagaskar. Den diskuterar även var man ska sätta gränsen mellan ett språk och ett annat. Över 400 olika språk berörs på ett eller annat sätt i boken, från abkhaziska till älv-dalska. Med hjälp av den senaste språkforskningen, men också av arkeologin och genetiken, pekar författaren ut både kända språk och språk som få hört talas om.

**Jonsson, Carla, Olvegård, Lotta & Tingsell, Sofia, 2021: Flerspråkighet i arbete. En vägledning från Språkrådet.** Uppsala: Institutet för språk och folkminnen. Publicerad på <https://www.isof.se/stod-och-sprakrad/vagledning-ar/flersprakighet-i-arbete>. 70 s.

Detta är en vägledning till alla som arbetar på en flerspråkig arbetsplats. Bokens första del är en introduktion till temat flerspråkighet, andra delen handlar om språkinläring och tredje delen tar upp kommunikation mellan förstaspråkstalare och inlärare. I slutet av vägledningen presenteras en lista med handfasta råd till förstaspråkstalare och arbetsgivare, sådant som man kan tänka på i samtal med andraspråkstalare. Frågorna i slutet av varje avsnitt är tänkta som underlag för den som leder språkarbetet på en arbetsplats, exempelvis när syftet är att ta fram en språkpolicy.

**Jonsson, Mats, 2021: När vi var samer.** Stockholm: Galago. 344 s.

I boken fortsätter författaren sitt självbiografiska serieprojekt med syftet att utforska det samiska ursprunget. Boken är både en släktkrönika och en bred, folkbildande skildring av det samiska folkets historia, från att Sápmi växer fram i Nordens norra delar till att svenskarna koloniserar området söderifrån.

**Jonsson, Åke, 2021: Radion och språket: om lyssnare, hallåmän, språkvård, dialekter och svordomar.** Första upplagan. Stockholm: Morfem. 208 s.

I boken berättar författaren sin egen och radions historia från ett språkligt perspektiv. Alltsedan Radiotjänst startade sina sändningar i januari 1925 har radions uppgift varit att förmedla nyheter, vara folkbildande och underhållande. Radion har samtidigt kommit att bli en självklar och naturlig plats för spridning av det talade ordet: såväl den föredömliga svenskan, som det så kallade språkliga fördärvet. Dialekter, svordomar, uttal av namn och ord i radio har diskuterats och debatterats från första stund, liksom frågan om vilka röster som får höras.

**Klarspråk, 2021.** Stockholm: Språkrådet, Institutet för språk och folkminnen. 4 s.

Klarspråk är en bulletin med nyheter om klarspråksarbete i Sverige och i andra länder. Tidningen kommer ut fyra gånger per år, och samtliga nummer går att läsa och ladda ned på Isofs webbplats: <https://www.isof.se/lar-dig-mer/publikationer/tidskrifter/bulletinen-klarsprak>. Man kan också prenumrera kostnadsfritt på tidningen. Nr 1 hade temat fackspråk och terminologi, där bland annat Erik Molander tog upp strategier för att lära sig ett yrkesspråk när svenska inte är ens förstaspråk. I nr 2 presenterades vinnaren av Klarspråkskristallen, nämligen Skatteverket som fick pris för sitt arbete med klarspråk under coronakrisen. Nr 3 hade temat demokrati, med anledning av att vi för 100 år sedan hade vårt första riksdagsval med allmän och lika rösträtt, och att kvinnor för första gången kunde ta plats i riksdagens kammare. I nr 4 på temat internkommunikation skrev bland annat Catharina Nyström Höög om mallar som ett hjälpmedel i klarspråksarbete.

**Larsson, Maria, 2021: Viktig läsning?: svensklärare i gymnasiet om läsundervisning i en performativ kontext.** Doktorsavhandling, Falun: Högskolan Dalarna. 311 s.

I avhandlingen undersöks hur svensklärare på gymnasiet talar om läsundervisningen i en kurs som avslutas med ett nationellt läsförståelseprov. Syftet är att bidra med kunskap om vilka didaktiska utmaningar svensklärare på gymnasiet ställs inför när de ska välja innehåll till läsundervisning i en i övrigt policybemängd och performativ kontext. Ett särskilt fokus riktas mot vad lärarna anser vara viktigt att elever lär sig och vilken *powerful knowledge* som synliggörs i den berättade praktiken.

**Lundblad, Carl-Erik & Josefsson, Gunlög, 2021: Småländska: orden och grammatiken.** Första upplagan Stockholm: Morfem. 332 s.

I boken används småländska som ett samlingsnamn för den språkliga bredden i Småland. Boken börjar med ett avsnitt om dialekter, och fortsätter med att beskriva småländskt uttal och ordbildning. Ordboksdelen tar upp småländska ord och uttryck tillsammans med språkexempel och förklaringar. Grammatikdelen beskriver och exemplifierar uttrycksmått som man finner i just småländska. Utgångspunkten är en äldre småländska som talades då Småland och resten av Sverige huvudsakligen beboddes av bönder, men här finns också inslag av den samtida småländska som man finner i litteratur och medier.

**Norrby, Catrin, Lindström, Jan, Nilsson, Jenny & Wide, Camilla, 2021: Interaktion och variation i pluricentriska språk: kommunikativa mönster i sverigesvenska och finlandssvenska: slutrapport från ett forskningsprogram.** Stockholm: Riksbankens jubileumsfond i samarbete med. 159 s.

Boken är en slutrapport över de viktigaste resultaten från forskningsprogrammet "Interaktion och variation i pluricentriska språk. Kommunikativa mönster i sverigesvenska och finlandssvenska" finansierat av Riksbankens jubileumsfond 2013–2020. I bokens första del redovisas programmets utgångspunkter, tidigare forskning om pluricentriska språk och samtalsmaterialet från tre domäner: service, lärande och vård. I dess andra del redovisas resultaten från olika interaktionella analyser, såsom hälsningsbeteende, samtalsöppningar, tilltalspraktiker m.fl. I bokens tredje del diskuteras resultaten i ett vidare sammanhang där mikroanalyser kopplas till kontextuella faktorer. Slutligen tolkas resultaten i en större samhällskontext och diskuteras med utgångspunkt i de relationer som råder mellan språkbruk, kultur och nation,

**Rösare, Sara & Mattson, Anki, 2021: Uppdrag textgranskning: en bok om att förbättra andras texter.** Andra upplagan, Stockholm: Producta. 170 s.

Uppdrag textgranskning är en praktisk handbok för dig som arbetar med att förbättra andras texter – med allt från att korrekturläsa till att göra större förändringar i språket och dispositionen. Boken tar på ett systematiskt sätt upp vad du behöver ta ställning till, oavsett om du kommer in i en verksamhet som konsult eller om texten är en kollegas eller en students.

**Spetz, Jennie, 2021: Nationella minoriteter. Vad vet svenska folket om dem?** Rapport. Uppsala: Institutet för språk och folkminnen. Publicerad på <https://www.isof.se/lar-dig-mer/publikationer/publikationer/2021-02-25-nationella-minoriteter-2020>. 36 s.

Institutet för språk och folkminnen, Länsstyrelsen i Stockholms län och

Sametinget har med hjälp av SifoKantar undersökt allmänhetens kunskap om och attityd till de nationella minoriteterna och bevarandet av de nationella minoriteternas språk och kulturer. Undersökningen gjordes i oktober 2020 genom telefonintervjuer med 1 000 personer. Respondenterna bestod av ett riksrepresentativt urval av befolkningen i Sverige (äldre än 15 år). Liknande undersökningar har gjorts 2010 och 2015 med det långsiktiga syftet att följa befolkningens kunskaper och attityder över tid. En utgångspunkt är att allmänhetens kännedom om de nationella minoriteterna kan ge en indikation på hur synliga minoriteternas språk och kultur är i samhället generellt.

### **Språktidningen, 2021.** 84 s.

Under 2021 har åtta nummer av Språktidningen kommit ut. I varje nummer finns en stor bredd av populärvetenskapliga artiklar, krönikor och reportage om både svenska språket och andra språk. Tidningen innehåller bokrecensioner och läsarnas frågor om språkbruk och dialekter med svar av språkvårdare, forskare och terminologer.

### **Strandberg, Mathias m.fl. (red.), 2020: Svenska landsmål och svenskt folkliv: tidskrift Årg. 143.** Uppsala: Kungl. Gustav Adolfs akademien. Publicerad på <https://kgaa.bokorder.se/sv-se/download/f553965f-e95e-4c22-9031-11b282e8cd06>. 233 s.

*Svenska landsmål och svenskt folkliv* är en gammal kulturtidskrift, grundad 1878 av J.A. Lundell och sedan 1997 utgiven av Kungl. Gustav Adolfs Akademien för svensk folkkultur. Sedan 2016 ges tidskriften ut i samarbete med Institutet för språk och folkminnen. Denna årgång innehåller bland annat Coppélie Cocqs keynote-föreläsning vid SIEF2019 om *Digitala avtryck och narrativ spårbarhet*, Sofie Henricson och Marie Nelson om *Variations-specifika gester. Att hålla tummarna i Finland och Sverige*, och Yair Sapir om *Sveriges samiska språk. Revitalisering, utmaningar och möjliga lösningar*.

### **Strömquist, Siv, 2021: Konsten att tala och skriva.** Sjunde upplagan Malmö: Gleerups. 272 s.

Konsten att tala och skriva är en klassisk handbok för alla som vill utveckla sin kommunikativa förmåga. Den tar upp hela arbetsprocessen, från planering, till formulering och bearbetning. Den redogör för skrivregler och kroppsspråkets betydelse varvat med många exempel. Den sjunde upplagan är grundligt genomarbetad och anpassad till dagens tal- och skrivsituationer samt de senaste årens utveckling på språkets och skrivandets område.

**Williams, Henrik, 2021: Rökstenen och världens undergång.** Stockholm: Kaunitz-Olsson. 398 s.

På Rökstenen, som står nära sjön Tåkern i Östergötland, finns världens längsta och mest komplicerade runinskrift. Stenen har varit känd sedan 1600-talet, och trots att man ägnat över 150 år åt att studera den har det varit svårt att få fram en entydig tolkning av texten. I den här boken vidareutvecklar författaren och hans forskarteam tidigare analyser av Rökstenen samt presenterar en ny tolkning av monumentet. Med gedigen forskning i grunden i kombination med ett populärvetenskapligt tilltal erbjuder författaren i Rökstenen en bredare publik att ta del av runstenens mysterier.

**Winlund, Anna, 2021: Inte för räddhågsna: undervisning i grundläggande litteracitet och svenska som andraspråk på gymnasieskolans språkintröduktion.** Doktorsavhandling, Göteborg: Göteborgs universitet. 269 s.

Syftet med denna avhandling är att öka kunskapen om undervisning för nyanlända elever med begränsade erfarenheter av formell skolgång, samt att vidga kunskapen om hur man tillägnar sig grundläggande litteracitet på ett andraspråk. Många av ungdomarna i elevgruppen har flyktingbakgrund men det är en heterogen grupp där eleverna kan ha mycket olika förutsättningar att hantera förväntade praktiker och diskurser i skolkontexten, och även mycket olika förutsättningar att utveckla litteracitet. Studien är etnografisk och insamlade data är från gymnasieskolans språkintröduktionsprogram i ämnet svenska som andraspråk och i samhällsorienterande ämnen.

**Utrikes namnbok: svenska myndigheter, organisationer, titlar, EU-organ och länder på engelska, tyska, franska, spanska, finska och ryska, 2021:** 11:e reviderade upplagan. Stockholm: Utrikesdepartementet, Regeringskansliet. Publicerad på <https://www.regeringen.se/informationsmateriel/2021/01/utrikes-namnbok/>. 148 s.

Utrikes namnbok innehåller namn på svenska myndigheter och organisationer, EU-organ och länder, samt titlar inom svensk offentlig förvaltning på svenska, engelska, tyska, franska, spanska, finska och (i viss utsträckning) ryska. Som i tidigare upplagor finns materialet på svenska, engelska, tyska, franska och spanska, men denna 11:e upplaga är dessutom den första sedan 2007 där hela materialet tagits fram även på finska.

# Adresser till språknämnderna i Norden 2022

Dansk Sprognævn, [www.dsn.dk](http://www.dsn.dk)  
Adelgade 119 B  
DK-5400 Bogense  
Danmark

Málráðið, [www.malrad.fo](http://www.malrad.fo)  
Handan Á 5  
FO-100 Tórshavn  
Färöarna

Institutet för de inhemska språken, [www.sprakinstitutet.fi/sv](http://www.sprakinstitutet.fi/sv)  
Hagnäskajen 6  
FI-00530 Helsingfors  
Finland

Stofnun Árna Magnússonar í íslenskum fræðum, [www.arnastofnun.is](http://www.arnastofnun.is)  
Skrifstofa Árnagarði  
Við Suðurgötu  
IS-101 Reykjavík  
Island

Oqaasileriffik, [www.oqaasileriffik.gl](http://www.oqaasileriffik.gl)  
Postboks 980  
DK-3900 Nuuk  
Grönland

Språkrådet, [www.sprakradet.no](http://www.sprakradet.no)  
Observatoriegata 1 B  
0254 Oslo  
Norge

Språkrådet, [www.isof.se/sprakradet](http://www.isof.se/sprakradet)  
Box 20057  
SE-104 60 Stockholm  
Sverige

Sámi Giellagáldu, [www.giella.org](http://www.giella.org)  
Sajos  
FI-99 870 Anár  
Finland

Nordisk Ministerråd, [www.norden.org](http://www.norden.org)  
Nordens Hus  
Ved stranden 18  
DK-1061 København K  
Danmark

Nätverket för  
språknämnderna i Norden

Språk i Norden är en årsskrift som språknämnderna i Norden publicerar. Vartannat år bygger innehållet huvudsakligen på föredragen från det nordiska språkmötet och vartannat år består skriften enbart av litteraturöversikter från de nordiska länderna. Årets tema är språkteknologi.

Språk i Norden  
Institutet för de inhemska  
språken, [www.sprakinstitutet.fi/sv](http://www.sprakinstitutet.fi/sv)  
Hagnäskajen 6  
FI-00530 Helsingfors  
Finland  
E-post: [may.wikstrom@sprakinstitutet.fi](mailto:may.wikstrom@sprakinstitutet.fi)