

# 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 Sprogmøde 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 numera-  
lier), to korte tekster samt en fri samtale (spontantale). Alle disse tekster var konstrueret således, at de skulle elicitere 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, optagelsessessioner 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               | sigá, kassi   | s%i:ja, k%as:I     |
|               |                                      | ʃ     | -/+    | ʃ               | skip, trýss   | S%i:b, tr%UJS:     |
|               |                                      | z     | -/+    | z               | úrslit, mars  | "Yzli:d, m%az:     |
|               |                                      | h     | -      | h               | húgva         | h%Igva             |
|               | Affricates                           | tʃ    | -      | tʃ <sup>h</sup> | tjúkkur       | tS%YHg:Ur          |
|               |                                      | dʒ    | -/+    | dʒ              | geva, ikki    | dZ%e:va, "IHd:ZI   |
|               | Nasals                               | m     | -/+    | m               | gloymd, mamma | gl%OJmd, m%am:a    |
|               |                                      | ɱ     | -      | ɱ               | gloymt        | gl%OJMd            |
|               |                                      | n     | -/+    | n               | máni, tunn    | m%OA:nI, t%Un:     |
|               |                                      | ɲ     | -      | ɲ               | tunt          | t%Uxd              |
|               |                                      | ŋ     | -      | ŋ               | sangur        | s%aNgUr            |
|               |                                      | ɳ     | -      | ɳ               | rankur        | r%EXgUr            |
|               | Laterals                             | l     | -/+    | l               | lítil, ball   | l%UJ:dIl, b%al:    |
|               |                                      | ɭ     | -      | 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%EaJ: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%lsgUrln  |
| 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 sprogressourcer 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 hverdagssprog, 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 stavetkontrol 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 Lamhauge** 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 Lamhauge*



*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 málskifti: 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, Göbel, Hans & Wolck, Wolfgang: *Kontaktlinguistik / 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
- Knoolhuizen, 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