region Västerbotten

ÁLGGUOGÅHTIE - UMESAMER I SAMVERKAN
Att: Samuel Hellgren
SÅGGATAN 33 D
414 58 GÖTEBORG

Beslut om stöd

Stödmottagare:	ÁLGGUOGÅHTIE - UMESAMER I SAMVERKAN
Org. nr:	802452-4566
Projektnamn:	Online tools for Inari Sámi and Ume Sámi
Ärendeid:	20370161
Diarienummer:	REGAC - 349 - 2024

Beslut om stöd

Region Västerbotten beviljar ÁLGGUOGÅHTIE - UMESAMER I SAMVERKAN stöd för att genomföra projektet Online tools for Inari Sámi and Ume Sámi enligt ansökan inkommen 2024-09-29. Stödet uppgår till 25,00 % av faktiska kostnader och 25,00 % av total finansiering, dock med högst 1 000 188.

Beslutet har fattats med stöd av förordningen (2003:596) om bidrag för projektverksamhet inom den regionala tillväxtpolitiken och förordningen (2015:210) om statligt stöd för att regionalt främja små och medelstora företag.

För stödet gäller allmänna villkor enligt Bilaga 1

Beslutet kan inte överklagas.

Motivering till beslut

Region Västerbotten beviljar stöd av följande skäl:
Projektet bedöms följa intentionerna i RUS prioritering 4. Västerbotten – en region med hållbara livsmiljöer att bo, verka och leva i och mer specifikt delmål 4.3 Utveckla kultur och kulturarv.

Projektet uppfyller kraven fastställda i Region Västerbottens Agenda för hållbar finansiering och bidrar i hög grad till prioritetshöjande aspekterna Sammanhållen region samt Positiva hållbarhetseffekter.

Projektperiod

 2025-09-01 - 2028-08-31

Projektbeskrivning

Syfte: Projektets mål är att skapa digitala språkverktyg för umesamiska och enaresamiska. Detta är två språk med relativt få talare, men med aktiv revitaliseringsrörelse. Språken är i sig tämligen olika varandra, men samarbete kring de tekniska aspekterna och erfarenhetsutbyte genom hela arbetet blir positivt för båda föreningarna. Eftersom denna

ansökan berör delfinansiering för den umesamiska delen av projektet så är det situation och mål för umesamiskan som beskrivs nedan.

Det traditionella umesamiska språkområdet är stort, och sträcker sig främst genom skogslandet mellan Ångermanälven och Piteälven. I det som idag är Västerbottens län har umesamiska talats främst i Sorsele, Malå, Storuman, och Lycksele kommuner, samt i kustkommuner som både haft bofast samisk befolkning samt berörts av renskötsel bedriven av umesamisktalande. Umesamiskan har till stor del varit ett skogssamiskt språk, vilket också är värt att nämna i relation till detta projekt. Den skogssamiska kulturen har drabbats hårt av koloniserande politik och strukturer. Att stärka det umesamiska språket i Västerbotten är därför en viktig insats för stärkandet av skogssamisk kultur, kulturarv och näring i länet.

Idag finns väldigt få digitala verktyg för umesamiska, t ex ordbok, paradigmgenerator, korpus, etc. Att få till stånd dessa verktyg vore av stort värde för alla som vill lära sig, bevara och utveckla det umesamiska språket. Detta språk är en viktig del av Västerbottens läns kulturarv, ett kulturarv som kan bevaras och utvecklas på ett innovativt sätt med de digitala verktyg som är projektets mål. Ett innovativt bevarat språkligt kulturarv kan också ge positiva effekter för kulturskapare och entrepenörer i samiska miljöer i Västerbottens län. Inte minst för barn och unga, som ofta har lätt att ta till sig ny teknik och inkorporera den i liv och lärande.

Vad projektets aktiviteter ska leda till för målgruppen på kort sikt:
På kort sikt publiceras digitala språkverktyg som underlättar inlärning, användning och utveckling av de båda språken.

Vilka effekter ska projektet uppnå på lång sikt:
På lång sikt bidrar projektet till revitaliseringen av umesamiska och enaresamiska. I Norrbotten kan de långsiktiga effekterna förväntas vara fler aktiva brukare av det umesamiska språket i både civilsamhälle och näringsliv. Det kan ses som ett stärkande av möjligheterna att uttrycka och utveckla kulturskapande med koppling till den umesamiska delen av urfolket samerna i Norrbotten. Inte minst för unga umesamiska språkbrukare, som är vana vid digitala verktyg och därmed särskilt kan förväntas dra nytta av projektets mål.

Indikatorer:

Namn	Värde	Måttenhet	Kommentar
Antal framtagna kunskapsunderlag, strategier och program.	1	Stycken	Språk verktyg ska utvecklas

Tid- och aktivitetsplan

Aktivitet	Beskrivning	Startdatum - Slutdatum	Kostnad
'SápmiFlex': Enhancing Inari Saami and Ume Saami Language Learning with Online Tools	We will develop an online tool for Inari Saami and Ume Saami word inflection, integrating the intricate principles of dictionary orthography for Inari Saami. This tool will serve as the primary interface for language experts accessing the new language model. The development of dictionary orthography needs a profound	2025-09-01 - 2028-08-31	

Aktivitet	Beskrivning	Startdatum - Slutdatum	Kostnad
	comprehension of Inari Saamis complex inflectional grammar. Traditional sources, such as textbooks, employ a model where words are categorized into groups, dissected into components, and subjected to inflection based on the interplay of these components. Each word group adheres to distinct inflection rules. Building upon existing research, Fabrizio Brecciaroli's Master's thesis further refined these models, particularly in analyzing disyllabic nouns. The main developer will leverage this foundation to construct a quantitative model of inflectional theory using the R programming language. The Ume Saami language community will create the necessary inflection rules for Ume Saami. This model facilitates automated word inflection and linguistic analysis, unlocking novel avenues for linguistic research and application. The project lead assumes responsibility for project management and communications pertaining to this work package. This entails overseeing the progress of tool development, ensuring adherence to project timelines, and facilitating effective communication among team members and stakeholders. The project lead will publish regular updates of the project progress and major releases through the established channels throughout the year.		
'SápmiFlex': Technical Implementation with R and R Shiny	The implementation of the online tool will utilize R/R Shiny technology for several reasons. Firstly, R is a widely used programming language in linguistics projects, renowned for its capability to produce clean, well-documented and testable code. Leveraging a high-level language like R facilitates the relatively straightforward integration of the developed linguistic model for Inari Saami and Ume Saami words. Moreover, R offers valuable functionalities such as streamlined packaging and dependency management of code components, alongside standardized support for automated testing and code documentation. The use of R Shiny further enhances the project by providing an intuitive front-end development framework, enabling the creation of user-friendly and modern web interfaces accessible on all devices: computers, tablets and smartphones. The emphasis on producing high-quality code is paramount, ensuring that the tool remains maintainable over an extended period and can be seamlessly handed over for further development, even if the original developers are not involved in future iterations. The parametrization of the developed linguistic model is facilitated by standard tables in CSV (or Excel) format, eliminating the need for IT expertise to comprehend the model's configuration. This approach empowers linguistic experts to modify rule sets and parameters independently. Additionally, the transparency of parameter tables ensures that all aspects of the model are explicitly visible, avoiding the inclusion of 'hidden' special cases within the code.	2025-09-01 - 2028-08-31	54 849
'SápmiFlex': Implementing Additional Inflection Tools	The Inari Saami and Ume Saami languages boast a multitude of potential inflection forms, presenting a significant challenge for learners. Beyond the fundamental forms like noun cases and main verb conjugations, learners must grapple with additional complexities. These include crafting passive forms, deminutives, possessive forms, as well as navigating adjective comparatives, superlatives, and adverbs. Moreover, thanks to the generative nature of our linguistic model, we can extend inflection to proper nouns such as personal and geographical names. In this Activity we	2025-09-01 - 2028-08-31	27 457

Aktivitet	Beskrivning	Startdatum - Slutdatum	Kostnad
	will implement these additional inflection tools to 'SápmiFlex'.		
'SápmiFlex': Implementing Word Analyzer Tool	After finalizing and validating the parameter tables with linguistic experts, we will generate an extensive database of inflected words. By systematically applying inflection rules to each word in our corpus, we will create a comprehensive linguistic resource. This database will underpin the development of a 'Word Analyzer' online tool, designed to provide detailed insights into any word in any form. For example, users will be able to determine a words classification into specific groups and sub-groups and retrieve information about its inflection form. The key difference between the Word Analyzer tool and the inflection tool lies in their functionality: while the inflection tool starts with a word's basic form and generates its various inflected forms, the Word Analyzer tool accepts an already inflected word and identifies its form, offering users precise linguistic analysis.	2025-09-01 - 2028-08-31	13 728
'SápmiFlex': Implementing E-dictionary, Text Corpus Browser, and Proofreader	Inari Saami already benefits from E-dictionary, Text Corpus browser and a proofreader, all maintained by the University of Tromsø. However, Ume Saami currently lacks these tools, which we aim to develop as part of this project. We will build the Ume Saami e-dictionary and text corpus browser directly on the R/R Shiny backend established in other project activities. The proofreader tool will be developed using the Ume Saami word corpus and the comprehensive set of inflected forms generated during this project.	2025-09-01 - 2028-08-31	109 829
Linguistic work for Inari and Ume Sami	The inflection theory for Inari Saami and Ume Saami is highly complex. Inari Saami is in a relatively good position, with most word classes having established and standardized rules for word inflection. However, there are still exceptions, particularly with three-syllable words, where further research and the development of new rules are necessary. Additionally, more word classes need to be defined to account for all potential special cases. Ume Saami, while having somewhat simpler inflection rules, faces challenges as many of its word classes and their corresponding inflection rules are neither standardized nor fully parameterized. This work package focuses on the linguistic groundwork essential for the developed online tools to accommodate all possible word inflection scenarios. The work involves analyzing texts and ultimately consulting language experts to clarify special cases through examples. The findings will then need to be systematically structured and parameterized by linguistic experts so that computer algorithms can accurately reproduce the correct inflection outcomes. The core linguistic work for Inari Saami is spearheaded by the Inari Saami Language Association, while the Ume Saami Language Association Álgguogåhtie undertakes the corresponding efforts for Ume Saami. The parametrization of the computational inflection model is a collaborative endeavor, guided by the project lead. Throughout the project, regular updates on the linguistic work will be shared via our media channels. Upon completion, the outcomes of this work package will be disseminated through publications in relevant forums, including international conferences and journals.	2025-09-01 - 2028-08-31	
Gathering Rules for Word Classification, Consonant Gradation and	Inari Saami and Ume Saami words can be categorized into distinct groups, each following unique inflection patterns. However, the classification of some words into these groups remains unclear, and further research is necessary. Computational analysis offers a powerful tool for identifying classification patterns that are difficult	2025-09-01 - 2028-08-31	78 450

Aktivitet	Beskrivning	Startdatum - Slutdatum	Kostnad
Vowel Families	to detect manually. Additionally, there is limited understanding of how specific words use consonant gradation or vowel families. Advanced computational linguistic research will be essential to uncover these patterns and rules. In cases where patterns do not exist, it will be necessary to compile word lists that indicate the correct classification for these words.		
Parametrization and Testing of the Linguistic Model with 'SápmiFlex'	The developed linguistic model necessitates parameterization using CSV tables, which will serve as the foundation for the 'SápmiFlex' tool's word inflection capabilities across all possible forms. To validate the accuracy of inflected words, we will leverage a large word corpus provided by the University of Tromsø as test data for Inari Saami and word corpus gathered by the Ume Saami Language Association as test data for Ume Saami. However, due to the extensive size of the corpus, even testing a subset of words within each word class poses a significant workload for linguistic experts. Automation will play a crucial role in streamlining testing processes. By comparing inflected words to existing inflection data we can conduct automated testing to identify discrepancies. Upon completion, the parameter tables will serve as a rich scientific resource, comprehensively detailing all inflection cases within the Inari Saami language. These tables will subsequently inform the creation of teaching materials in more accessible formats, such as printed inflection tables, facilitating broader dissemination of linguistic knowledge.	2025-09-01 - 2028-08-31	98 062
Investigating Incomplete Inflection Cases	Despite years of research, numerous aspects of inflection remain inadequately understood. To develop a comprehensive inflection tool, it is imperative to address these knowledge gaps. While some gaps are acknowledged, others may not even be recognized. Automated linguistic tools can play a pivotal role in identifying areas warranting further investigation. For instance, if the linguistic model fails to accurately inflect certain words, it signals the need for linguistic experts to delve deeper into these areas. This may involve proposing new word classes and refining rules to delineate distinctions between classes.	2025-09-01 - 2028-08-31	49 031
Gathering word and text corpus for Ume Saami	Ume Saami currently lacks an openly accessible, centralized and well-curated resource for its words and texts. Inari Saami is in a better position and in this project we can use the word and text corpus resources already maintained by the University of Tromsø. While Ume Saami has existing resources, such as word lists, dictionary publications, texts and audio recordings, they require meticulous curation and organized storage. In particular, the audio recordings need to be transcribed, which is a labor-intensive process. The Ume Saami Language Association Álgguogåhtie will lead this work package, coordinating task distribution among various members of the language community and managing all communication efforts. The Inari Saami Language Association will provide guidance and support, leveraging its decades of practical experience in speech-to-text transcription and the development of word and text corpora. At the outset, the Ume Saami Language Association will engage the language community through their media channels to recruit participants for the project. They will also ensure consistent media visibility throughout the project and, upon completion, will publish the results in relevant channels, including	2025-09-01 - 2028-08-31	

Aktivitet	Beskrivning	Startdatum - Slutdatum	Kostnad
	scientific conferences, journals, and regular media outlets.		
Transcription of Ume Saami recordings	Since the early 1900s, various linguists and researchers have collected an abundance of language-related materials from Ume Saami speakers, including both written and audio records. However, scholars have recently pointed out that very little research has been conducted with these materials. It appears that earlier linguists were primarily focused on gathering data and did little to analyze or study their findings.

Currently, the Institutet för språk och folkminnen (Institute for Language and Folklore, abbreviated as ISOF) holds audio recordings of native Ume Saami speakers from across the entire language region, most of which were made between the 1940s and the 1980s. However, these recordings remain difficult to access for the average Ume Saami student or language user, creating a significant barrier. This is particularly problematic because these recordings of native speakers are invaluable for language revitalization efforts and further research.

A key part of the current project aims to transcribe a selection of these recordings, which will expand the available resources for online tools and improve accessibility for students and other language users. The Ume Saami Language Association will lead this initiative, delegating tasks to various members of the language community and overseeing communications. The Inari Saami Language Association will provide guidance and support throughout the process. | 2025-09-01 - 2028-08-31 | 183 561 |
| Gathering of Ume Sami word lists and word inflection tables | Ume Saami word lists are currently available in different places and forms. One important asset is the 2018 printed dictionary edited by Henrik Barruk. Various teachers and language users keep their own personal dictionaries, with words retrieved from older Ume Saami texts such as the 1780 Lexicon Lapponicum and a variety of other pre-modern sources that have recently been made available online, and also from audio recordings from ISOF (mentioned above). Álgguogåhtie recently published a printed book consisting of some of the more common verb inflections. Similar collections of inflections for verbs, nouns and adjectives are kept by Ume Saami language teachers. An aim of the present project would be collecting these separate word lists and inflection tables, with the consent of the teachers and language users currently holding them.
The Ume Saami Language Association Álgguogåhtie will lead this work package, distributing tasks among various members of the language community and being responsible for the communications. The Inari Saami Language Association will offer guidance and support. | 2025-09-01 - 2028-08-31 | 183 561 |
| Gathering of Ume Saami text corpus | Since the establishment of an official orthography for Ume Saami in 2016 a small but thematically varied corpus of Ume Saami texts has been published. The current composition of Ume Saami texts publically available consist of childrens books, news articles, forewords and abstracts to scientific papers and theses, poetry and yoik-texts, bureaucratic texts, a beginners educational textbook, Christian hymns, and other religious texts. The aim of this particular project activity would be gathering and processing these texts to assemble a digital text corpus, freely and easily searchable by any language user. Such a tool would contribute immensely to language proficiency among language users on all levels, since it would be possible to inquire the usage of words in different literary and social contexts. This activity would also require some effort to rectify some | 2025-09-01 - 2028-08-31 | 91 786 |

Aktivitet	Beskrivning	Startdatum - Slutdatum	Kostnad
	of the older published texts to ensure they are up to code regarding the orthography, as some aspects of the orthography have been clarified and corrected since the publication of some of the Ume Saami texts. The Ume Saami Language Association Álgguogåhtie will lead this work package, distributing tasks among various members of the language community and being responsible for the communications. The Inari Saami Language Association will offer guidance and support..		
Creation of teaching materials	The word lists, text corpus and parameter tables we've developed represent a rich linguistic asset. To enhance accessibility, we plan to craft teaching materials derived from these findings. By condensing complex inflection rules into human-readable tables, we aim to provide invaluable resources for quick reference on word inflection. While suitable for all education levels, these materials are particularly tailored for higher education settings like high schools and universities. In addition, we will write guide books to aid Inari Saami and Ume Saami writers. The work in this package will create essential resources to enable the teaching of Ume Saami in higher education and universities. Project management and communications for this work package are collaboratively handled by the Inari Saami Language Association and the Ume Saami Language Association Álgguogåhtie. The project lead will provide support to those involved, particularly in the editing and publication of printed materials. The developed materials will be widely promoted across our media platforms, and the printed resources will be prominently featured in relevant Saami bookstores and libraries.	2025-09-01 - 2028-08-31	
Creation and publication of inflection rule tables for Inari Saami and Ume Saami	During the project, we will develop detailed, information-rich CSV-based inflection tables primarily for use by our computer software and linguistic researchers. While these tables contain all the necessary information, they are not user-friendly for language learners due to their complexity. Therefore, we will simplify and condense the information into more accessible formats. To make these learner-friendly tables easier to review and understand, we will incorporate visual cues such as colors and text formatting. The tables will be published online in PDF format, making them easy to print. Printed versions are especially valuable for language learners and those less familiar with computers, such as the elderly, offering a convenient reference for inflection rules across various word classes. They will also serve as a useful tool in schools, supporting language teaching.	2025-09-01 - 2028-08-31	14 121
Creation and Publication of writers guide in Inari Saami and Ume Saami	The Inari Saami and Ume Saami literary communities currently lack comprehensive resources to support writers. Much of the essential grammatical and syntactic information is scattered across various sources, making it difficult for writers to find what they need. As part of this project, we will compile all relevant information into cohesive resource books. Since Inari Saami and Ume Saami are both part of the Finno-Ugric language family, we can efficiently collaborate to create writers' guides for both languages at the same time. While the guides will share a similar structure, we will carefully address the specific differences between Inari Saami and Ume Saami. To make them more accessible for learners not yet fluent in the languages, the	2025-09-01 - 2028-08-31	56 484

Aktivitet	Beskrivning	Startdatum - Slutdatum	Kostnad
	Inari Saami guide will be written in Finnish, and the Ume Saami guide in Swedish. The resource will cover a wide range of topics, including interference from Finnish and Swedish, verb government, the use of case and dual forms, possessive suffixes, word formation, pronoun usage, and numerous examples of word usage. It will also address additional grammar topics not covered by existing grammar books and include exercises to help correct poorly written phrases. These guides will be provided in accessible PDF formats, with printed copies available as well. They will serve as a valuable resource for Inari Saami and Ume Saami learners and writers, ensuring consistent and accurate language use. Additionally, they can be used as reference material in schools and universities.		
Yearly Saami language technology conference	We will organize a yearly Saami language technology conference. The participation is not limited to project members; all interested individuals are welcome to attend. Additionally, the sessions will be available for online viewing. Project management and communications for this work package are collaboratively managed by the Inari Saami Language Association and the Ume Saami Language Association Álgguogåhtie. The communication efforts will include promoting the event to the broader Saami and linguistic communities across Europe and globally. This will be achieved through appropriate channels, such as scientific online forums, and proactive outreach to our associate partners and contacts well in advance of the event.	2025-09-01 - 2028-08-31	
Language technology conference organization	The yearly language technology conference will cover a variety of topics. The agenda includes updates on the project's progress and presentations by invited speakers who will showcase various language technology tools, extending beyond just Inari Saami and Ume Saami. Additionally, speakers will share their experiences in content creation on platforms like Wikipedia and YouTube. The conference will also feature training sessions on using language technology tools, writing for Wikipedia, and participating in open-source projects. There will be in-depth under the hood presentations that explain how these language technology tools function. The organizational work for the conference is a collaborative effort each year. The event will alternate annually between Inari and Ume, with the local organizer in charge of managing logistics, including securing the venue.	2025-09-01 - 2028-08-31	19 612

Budget (Kostnads- och finansieringsplan)

Kostnad

Kostnadsslag	2025	2026	2027	2028					Totalt
Personal	317 523	955 623	955 623	637 082					**2 865 851**
Schablonkostnader	127 005	377 961	377 961	251 974					**1 134 901**
Summa kostnader	**444 528**	**1 333 584**	**1 333 584**	**889 056**					**4 000 752**
Projektintäkter									
Summa faktiska kostnader	**444 528**	**1 333 584**	**1 333 584**	**889 056**					**4 000 752**
Bidrag i annat än pengar									
Summa bidrag i annat än pengar									**0**
Summa totala kostnader	**444 528**	**1 333 584**	**1 333 584**	**889 056**					**4 000 752**

Finansiering

Finansiär	2025	2026	2027	2028					Totalt
Offentligt bidrag i annat än pengar									
Total offentligt bidrag annat än pengar									**0**
Offentlig kontantfinansiering									
LÄNSSTYRELSEN I NORRBOTTENS LÄN	288 943	866 830	866 830	577 886					**2 600 489**
REGION NORRBOTTEN	44 453	133 358	133 358	88 906					**400 075**
Total offentlig kontantfinansiering	**333 396**	**1 000 188**	**1 000 188**	**666 792**					**3 000 564**
Total offentlig finansiering	**333 396**	**1 000 188**	**1 000 188**	**666 792**					**3 000 564**
Privata bidrag i annat än pengar									
Total privat bidrag annat än pengar									**0**
Privat kontantfinansiering									
Total privat kontantfinansiering									**0**
Total privat finansiering									**0**

Stöd

Finansiering	2025	2026	2027	2028					Totalt
19.1.1 Regionala utvecklingsåtgärder Regionalt projekt	111 132	333 396	333 396	222 264					**1 000 188**

Sammanställning (Stödprocent)

Stödandel av faktiska kostnader: 25,00 %

Stödandel av stödgrundande finansiering: 25,00 %

Stödandel av total finansiering: 25,00 %

Andel annan offentlig finansiering: 75,00 %

Andel privat finansiering: 0,00 %

Rapportering och begäran om utbetalning
Stödet utbetalas i efterhand efter redovisning av faktiska utgifter.
AoU synkas med Aurora programmet.

Sista datum för slutrapport
2028-10-31

Allmänna villkor för stöd

Se bilaga

Särskilda villkor

--

Beslut i detta ärende har fattats av Regional utvecklingsdirektör, Patrik Sällström efter föredragning av Mayuri Kumari.

Vid frågor kontakta:

Mayuri Kumari
Telefon: 0736909860
E-post: mayuri.kumari@regionvasterbotten.se