



Language Technology Programme for Icelandic 2024-2026



The Government Offices of Iceland
The Ministry of Culture and Business Affairs

Prepared by the steering group to develop the next Language Technology Programme for Icelandic, appointed by the Minister of Culture and Business Affairs in May 2023. The group consisted of chairman Björgvin Ingi Ólafsson, Lilja Þógg Jónsdóttir and Páll Ásgeir Guðmundsson. Óttar Kolbeinsson Proppé collaborated with the group on behalf of the Ministry of Culture and Business Affairs.

Responsible party: The Ministry of Culture and Business Affairs

Sölvhólsgrata 7 – 101 Reykjavík
545 9800 | mvf@mvf.is

Published by: The Government Offices of Iceland

Editing: The Government Offices of Iceland

Layout: Einar Guðmundsson

Proofreading: Óttar Kolbeinsson Proppé

Photos, unless otherwise specified:

Unsplash and Freepik

March 2024

ISBN: 978-9935-9743-5-8

© 2024 – The Ministry of Culture and Business Affairs

Table of Contents

Summary	5
1. Introduction	9
1.1 Scope	9
1.2 Background	11
1.3 Iceland's responsibility and policy regarding the Icelandic language	13
1.4 Advanced language technology for general utilisation	16
1.5 Extensive utilisation	16
1.6 Continued international activities	17
2. Operation of the Language Technology Programme	21
2.1 Artificial intelligence and language technology	22
2.2 An Icelandic LT and AI centre	23
2.3 Administration in the coming years	26
2.4 Funding arrangements for language technology	27
3. Utilisation of language technology	31
3.1 Funding for utilisation projects	31
3.2 Counselling for Icelandic industries and general introduction of LT solutions	32
3.3. Communication with foreign tech companies	33
4. Core Projects	37
4.1 Programme flexibility and new arrangements	37
4.2 Core projects and competition considerations	40
4.3 Prioritised core projects	40
4.4 Maintenance of infrastructure and core solutions	48

Summary

The Icelandic government's previous Language Technology Programme, completed in 2023, provided an ambitious vision for the preservation of the Icelandic language and made Iceland and Icelandic better equipped than most other small linguistic areas to cope with rapid technological changes in artificial intelligence (AI) and language technology (LT).

The charter of Iceland's current government states that efforts to strengthen the position of Icelandic in the digital world will continue, with a special focus on language technology. To that end, in May 2023, the Minister of Culture and Business affairs appointed a steering group to evaluate and analyse the best methods for long-term integration of language technology at the administrative level, as well as to draft the next official Language Technology Programme.

The steering group initially worked with a working group on language technology appointed by the Minister of Culture and Business Affairs alongside the steering group. The working group was made up of 12 representatives of language technology companies, universities, institutions, and organisations of those who can benefit from language technology in their daily activities, as well as members of the steering group. The group held four meetings during its term of appointment where it succeeded in highlighting the important views of stakeholders and their priorities for the next Language Technology Programme. In addition, the project period of the previous Language Technology Programme was dissected with a view to learning from experience and identifying opportunities for improvement when making the next steps for Icelandic language technology. This proved to be highly useful preparation for developing the proposals presented herein.

The report covers three main topics:

Maintenance and continued development of infrastructure developed during the project period of the previous Language Technology Programme.

The utilisation and implementation of language technology in Icelandic industries, government and society. There is particular emphasis on the **promotion of Icelandic language technology** with the aim of bringing Icelandic language technology solutions to software development, both in Iceland and to foreign tech companies.

Integrating AI and language technology issues to strengthen Iceland's competitive position and increase prosperity in society, with language at the forefront.



The report proposes the allocation of funds for language technology in 2024-2026 to achieve the objectives and priorities proposed by the steering group. It is also proposed that work be initiated to establish a future common framework in Icelandic administration for language technology and artificial intelligence.

It is important that an overall strategy for language technology and artificial intelligence be formulated by the end of 2024, and ideally such a strategy should be adopted by the Icelandic Parliament.

A position must be taken on whether to establish a language technology and AI institution according to the Nordic model, and how such an institution could be funded. It is also necessary to assess how the areas of artificial intelligence and language technology can be better integrated at the administrative level.

The steering group proposes that the Ministry of Culture and Business Affairs collaborate with the Ministry of Higher Education, Science and Innovation to initiate work on such policies through a feasibility study on a new language technology and AI institute. It is important to find the most suitable venues for these issues, and for the period 2024-2026, it is proposed that the Ministry of Culture and Business Affairs undertake the following:

-
- 1** **LT and AI centre:** The Ministry of Culture and Business Affairs will collaborate with the Ministry of Higher Education, Science and Innovation to explore the feasibility of establishing a new LT and AI institute based on Nordic models. The feasibility study will be carried out in 2024, following which the ministries will work together to submit to Parliament a policy on artificial intelligence and language technology for the next ten years.

 - 2** **Utilisation grants:** ISK 60 million will be spent annually on utilisation projects in language technology. These grants should not only support the development of end solutions that make use of language technology infrastructure developed under governmental language technology programmes, but also support the implementation of these solutions in the systems and services of companies.

 - 3** **Promotion and consultancy:** A much greater emphasis will be placed on promotion and consultancy activities for language technology in Iceland. There will be much more promotion of the infrastructure developed for Icelandic language technology, and the relevant opportunities for Icelandic industries, non-governmental organisations and the public. We will also systematically monitor the technological developments of foreign tech companies and make efforts to introduce Icelandic language technology to them.

 - 4** **Core projects:** Between 2024 and 2026, at least ISK 100 million will be allocated annually for continued development of core projects in language technology. The projects proposed in this report will be re-evaluated annually, under the plan's new project management arrangement, in accordance with the rapid pace of technological developments.

 - 5** **New arrangement of maintenance:** A new arrangement of maintenance for language technology infrastructure will be established and ISK 40 million allocated for maintenance each year up to and including 2026. The project manager will oversee the arrangement and allocate funding to projects as needed in cooperation with the project management.

 - 6** **The government's leading role in the implementation of language technology:** The government will be at the forefront of implementing language technology solutions in its services, both within ministries and their agencies.

 - 7** **The CLARIN project:** Iceland's participation in the CLARIN partnership, a EU-funded research infrastructure project, will be re-evaluated and the scope and costs of the partnership, which is currently hosted by the Árni Magnússon Institute under a 2026 agreement, will be reviewed. The evaluation would be carried out in parallel with the Ministry's LT and AI policy development following the conclusion of this project plan in 2026.
-

Chapter 1



1. Introduction

1.1 Scope

With a mandate dated 15 May 2023, Björgvin Ingi Ólafsson, as chairman, Lilja Dögg Jónsdóttir and Páll Ásgeir Guðmundsson were appointed to the steering group to draw up the next Language Technology Programme. The group was required to submit proposals to the Minister of Culture and Business Affairs on arrangements for the implementation of language technology projects, taking into account the role of the government and prioritising language technology projects for the period 2023-2026. The main objectives are to ensure that there is continued progress in the development of Icelandic language technology solutions after the end of the project period of the previous Language Technology Programme, *Language Technology for Icelandic 2018-2022*, as well as to ensure that the infrastructures developed under this programme are put to increased and widespread use by the public and in the economy. This places greater emphasis on the practical utilisation of language technology.

The steering group was subsequently also asked to review the administrative arrangements for the previous Language Technology Programme and to propose a longer-term arrangement, as well as to consider in particular how to ensure efficient use of the funds provided for in the 2023 budget.

In addition to appointing the steering group, the Minister appointed a working group on Icelandic language technology to advise and consult the steering group as needed at the beginning of the process, and as a consultation forum for the reporting of the steering group. The working group was composed of representatives of 12 stakeholders; language technology companies, universities, governmental ministries, public institutions and NGOs, as well as the three members of the steering group. Björgvin Ingi Ólafsson was also appointed chairman of the working group.

The working group held four meetings in the period May-August 2023, and questionnaires were sent out to all members of the group, asking for input on various factors that were to be addressed. In addition to this, the steering group met numerous times with stakeholders, foreign experts and governmental representatives. These meetings were successful in highlighting important stakeholder perspectives for the next Language Technology Programme. The project period of the previous Language Technology Programme was analysed with a view to identifying what could have been improved and where the main opportunities for improvement lay for the next phase of language technology work in Iceland.

The working group found that the perspectives were multifaceted and there were various issues to be resolved. Stakeholders differed somewhat on various focus points and on the most sensible ways to formulate future arrangements for Icelandic language technology. It also bears mentioning that the stakeholders appointed to the working group had varying levels of previous involvement in language technology. This mostly applied to the representatives of the ministries in the group.

Main challenges:

1

The venues for language technology and AI

The venues for AI and its derived function, language technology, must be intertwined.

2

Clear division of labour

The integration of technical development, utilisation, promotion and communication with foreign companies and governments is essential. The roles of the different parties need to be better defined than was the case in the previous programme. Different factors require different competences and it is important to ensure an overview of multiple tasks under a single project management responsible for the implementation of the Language Technology Programme.

3

Language technology at the administrative level

Knowledge of language technology must be promoted within the administrative level. The willingness and mandate of the ministries to support or work systematically towards the implementation of language technology solutions in public services is not explicit enough.

4

New technologies

There are differing perspectives on the possibilities of AI and large language models and the difference that such technologies might make for Icelandic language technology. There are also differences of opinion on the need for continued development of domestic solutions built from scratch when adapting technology developed abroad could be more effective and cost-efficient.

5

Competition guaranteed

The importance of competition considerations and that the public sector does not compete with domestic tech companies through the development of core projects which are entirely publicly funded. At the same time, it is necessary to define how far the concept of a core project extends and when utilisation in the economy takes over.

6

Maintenance of infrastructure

Maintenance of the language technology infrastructure that has been developed must be ensured, and it must be determined which solutions are essential for the government to maintain.

1.2 Background

The Icelandic government's Language Technology Programme published in June 2017, the *Language Technology for Icelandic 2018-2022 Project Plan*, was a milestone in the development of language technology in Iceland. The programme brought about investments in language technology infrastructure through generous public funding and concentrated efforts of specialists under the banner of the SÍM Consortium. The SÍM Consortium was a group of Icelandic universities, institutions, companies and non-governmental organisations that worked on research and development under the programme pursuant to agreements with the private non-profit institution Almannarómur. SÍM worked on the development of core projects in language technology, while Almannarómur was responsible for the project management on behalf of the Ministry. The Language Technology Programme identified 6 core projects, which have laid the foundations for the development of language technology solutions for Icelandic. These were core projects in the following categories, published in open source repositories:

Language resources

All language technology is built on language resources: texts and/or audio recordings. These are necessary in the analysis of language, identification of new and emerging words and senses, and in finding rules and patterns. From language resources it is possible to "teach" the computers what is important for the software being developed, or use software to identify rules and patterns in large volumes of data. Within the Language Technology Programme, large text collections have been developed and prepared for use in language technology, both monolingual Icelandic text collections and bilingual parallel corpora containing Icelandic and English texts. Large-scale voice recordings are also underway, both through the crowdfunding project Samrómur and high-quality studio recordings used to develop speech synthesisers. Databases containing information on individual aspects of the language, such as vocabulary, pronunciations and semantics, have also been developed.

Spell and grammar checking

Spell and grammar checking assists with correcting text and using correct spelling, even in the appropriate register. Spell and grammar checking is also important for the development of other language technology software, as errors in the written text can affect its further automatic processing. The aim of the Language Technology Programme has been to develop a general spell and grammar checker that can detect and correct the most common errors made. The aim is to gather knowledge about the nature of typographical errors in different groups and to develop methods to adapt the system to different needs, including in the fields of training and teaching.

Speech synthesis

Speech synthesis converts written text into speech. Two main areas of speech synthesis software are dictation and (verbal) communication. Speech synthesisers are used to read text, for example from websites or books. People who for some reason are unable to read, or struggle with reading, rely on speech synthesis in their daily lives. Communication systems, in which a speech recogniser hears what the user is saying, need speech synthesisers to be able to answer in a human-like voice. The Language Technology Programme has focused on developing new speech synthesiser voices for Icelandic, for reasons that include allowing users to choose a voice that they find comfortable.

Speech recognition

Speech recognition is about converting speech into written text. It is the prerequisite for us being able to communicate with computers and machines in a way that is natural to most of us – by speech. The objective of the Language Technology Programme has been the creation of a general Icelandic speech recogniser that can be used with web services. All methods and data will also be available as a basis for the development of specialised speech recognisers.

Machine translation

Machine translation (MT) is the automatic translation from one language to another. This has already become useful for many language pairings, to help people understand the content of text that is written in a language they are not familiar with, and to expedite the work of translators in languages in which they are fluent. As of yet, however, no translation software is capable of delivering translations that are of fully satisfactory quality. The text must always be reviewed and corrected to provide an accurate translation. The objective of the Language Technology Programme has been to create an open translation engine that translates between Icelandic and English. Such an engine should be useful for translations in particular domains, enabling translators to complete their translations faster than before.

1.3 Iceland's responsibility and policy regarding the Icelandic language

Icelandic language objectives

The charter of Iceland's current government specifically addresses the rapid technological developments and digital transformation taking place all across the globe, adding that during its present term, the government would provide further support for technical infrastructure, as well as focusing on innovation and the utilisation of inventive solutions.

To quote the charter:



Efforts to strengthen the position of Icelandic in the digital world will also continue, with a special focus on language technology.

The charter also contains a pledge to pursue utilisation of new technologies and AI, with particular support for transformation in the educational system, but also in health and biosciences, software development and green technology, and the creative industries.

The Act on the Status of the Icelandic language and Icelandic Sign Language no. 61/2011 furthermore states that:



The state and local governments have a responsibility to preserve and strengthen the Icelandic language and shall make sure that it is used.

As of the writing of this report, a proposal for a parliamentary resolution on an Icelandic language action plan for the years 2023-2026 has been submitted in Parliament. This proposal is the result of a collaboration between the five ministries represented in the Council of Ministers for the Icelandic Language. Action 16. *The future of language technology* stipulates that



...work will be undertaken to develop a new language technology programme that will reflect the government's priorities for the continued development, maintenance and implementation of language technology solutions. Among the focus points in language technology projects up to 2026 would be solutions and projects that are useful to the public.

Icelandic Language Policy

In September 2021, the Icelandic Language Council issued a new Icelandic Language Policy for 2021-2030. The policy specifically addresses language technology and reiterates its importance for the future of the Icelandic language:



Keeping up with developments in language technology is essential if Icelandic is to remain an official language and the national language of Icelanders. Such a project must not be a short-term effort but must rather be seen as a never-ending project that is constantly evolving in line with the rapid technological developments.

The Language Policy also calls for the government to implement a Language Technology Programme, support development projects and start-up companies in language technology, promote Icelandic language technology and encourage Icelanders to take advantage of such technology.

The Fourth Industrial Revolution and AI Policy

In mid-2018, the Prime Minister appointed a committee on the Fourth Industrial Revolution to “report on global discourse regarding the Fourth Industrial Revolution, its implications for Icelandic society and opportunities that these changes may provide for Iceland”. A common theme throughout the committee’s conclusion is the importance of Icelandic society embracing technological developments in the understanding that the effects of technology depend on how people choose to use it. The committee also took pains to stress that increased productivity due to automation must be fairly distributed throughout society.

The use of Icelandic in new technologies is precisely a key prerequisite for this as can clearly be seen in the action plan developed on the basis of the committee’s report by a project management board appointed by the Prime Minister in July 2019. Two of the 27 actions detailed in the action plan specifically address Icelandic language technology:



[Action 26] It must be ensured that Icelandic is viable in a digital world. The project management proposes to draw up a plan of what will follow once the 5-year Language Technology Programme has concluded. [...]

[Action 27] Efforts will be made to create an automatic translation system between English and Icelandic for faster and less costly Icelandic subtitling. An automated translation system would also be useful for digital content providers.

Another action in the plan concerned the importance of Iceland establishing a policy on artificial intelligence, which had not been done before. In October 2020, the Prime Minister appointed a committee to draw up an AI policy. In April 2021, the committee submitted its suggestions, which include a special discussion of the Icelandic language and the importance of language technology research to ensure that AI can process data in Icelandic.

In preparing this report, the above policies and parliamentary resolutions were taken into account and the proposals presented herein are consistent with the objectives of the government and the resolutions of the Parliament, both in the field of Icelandic language and AI.



1.4 Advanced language technology for general utilisation

The Icelandic Government's Language Technology Programme, completed in 2023, was an ambitious vision for the protection of the Icelandic language and made Iceland and Icelandic better equipped than most other small linguistic areas to cope with this transition.

The Language Technology Programme not only guaranteed the development of necessary infrastructure to support the Icelandic language in a digital world but also worked to ensure that Icelandic would forever remain our language.

During the project period of the previous Language Technology Programme, *Language Technology for Icelandic 2018-2022*, dozens of specialists collaborated on developing open core solutions, led by the SÍM Consortium. In retrospect, the programme was a success and marked a remarkable step forward in the development of language technology, proving how smaller languages can have an impact in a digital world.

In the phase of the Language Technology Programme presented here, there is a strong emphasis on how to ensure the most efficient use of the infrastructures that have been developed, at companies, institutions and municipalities in Iceland and among international companies, where it is important for Icelandic to be one of the languages included in technological solutions. In addition to this, there is a focus on maintenance and continued development of the infrastructure already developed.

Policy-making in technologically-related fields can pose various difficulties, as developments tend to be rapid and unpredictable. In recent years, the importance and potential of large language models have far exceeded what was assumed in the previous Language Technology Programme. We can expect to see continued technical advancements, but the nature of future developments is as unclear as always, although it can be predicted that multi-faceted language technology problems can be solved on the basis of large language models, and in a more comprehensive manner than was previously possible. This is all highly uncertain, however. These developments underline the importance of establishing policies that are both forward-looking and unequivocal but not set in stone, allowing them to be amended in line with circumstances at each time.

1.5 Extensive utilisation

It is important that the next Language Technology Programme focus on the utilisation of existing language technology infrastructure. Thus, the new Language Technology Programme focuses more on the utilisation of language technology in a wide sense, along with the maintenance and development of language technology solutions. More emphasis will be placed on promoting current and future infrastructure and solutions to increase utilisation to the extent possible. There will be a particular focus on the

utilisation of language technology within the education and healthcare systems, as language technology and other AI-powered technological developments can clearly have various positive effects on both systems.

Great efforts will also be made to encourage business-driven innovation based on available language technology infrastructure. Start-up companies have already introduced language technology solutions based in part on available infrastructure, while larger companies have adopted a strategy to apply language technology and recruited experts to drive further development of language technology within the companies. It is a fact, however, that there are still many untapped opportunities for companies and organisations in implementing and utilising language technology solutions. It is important that the authorities contribute to advance further utilisation of language technology in the corporate sector, by providing both more effective information on the infrastructure and solutions in place, and direct support for projects to introduce and apply language technology as part of economic activities.

1.6 Continued international activities

In recent years, Iceland's language technology efforts have included presenting the infrastructures that have been developed to major foreign tech companies to point out the relative ease of including Icelandic as part of global solutions, due to the infrastructure that has been developed in Iceland. Presentations have been made for leading tech companies including Microsoft, Amazon, Apple, Google and Meta. Despite success in reaching many of these companies, more systematic work must be carried out to ensure that Icelandic will be among the languages offered. Recent examples include both solutions where Icelandic is in a prominent position and solutions where Icelandic did not gain access.

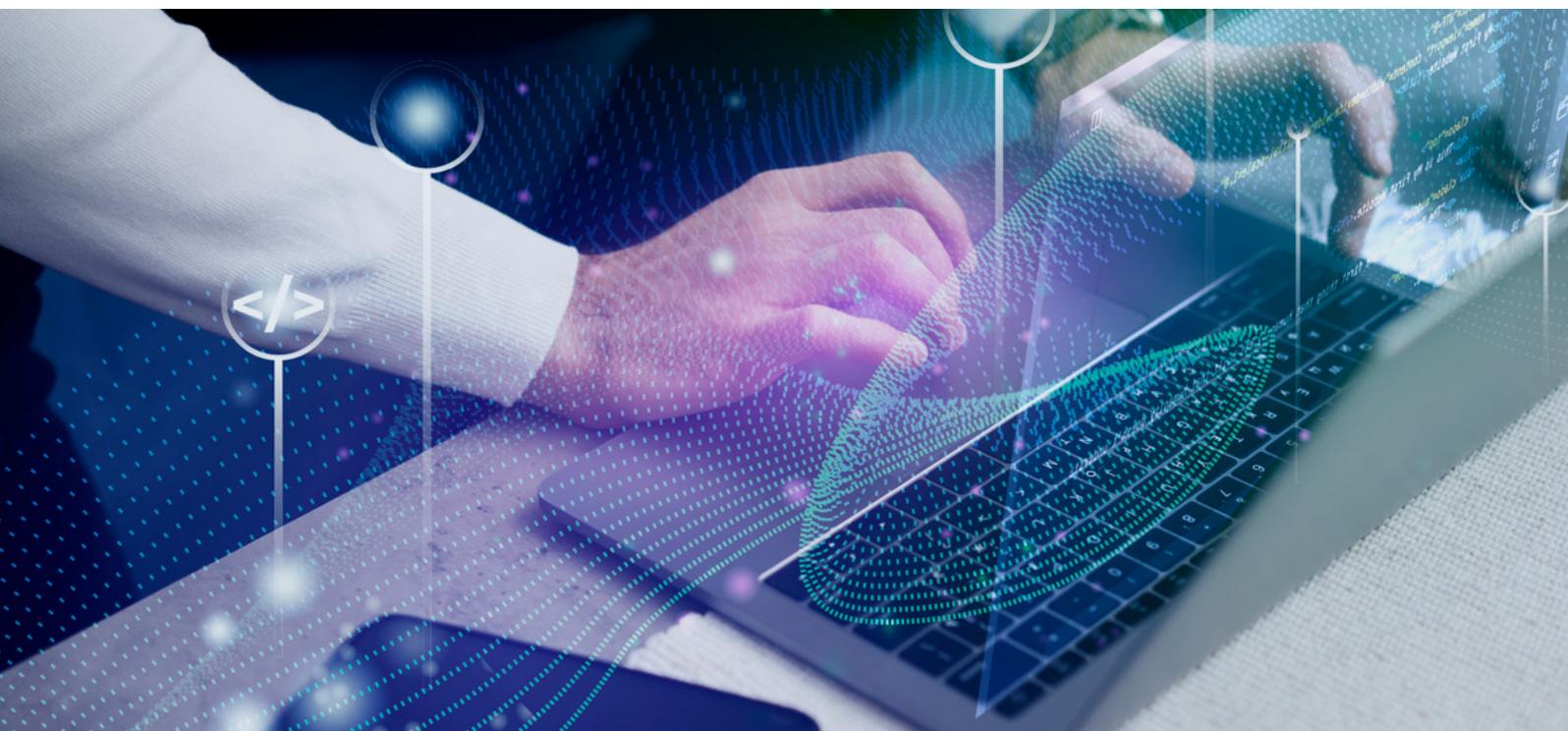
While introducing the products of the previous Language Technology Programme, a strong relationship formed with Open AI, the author of the language model Chat GPT. When Chat GPT-4 was launched in November 2023, a specially selected Icelandic dataset was incorporated into the language model. Thus, great progress was made for all Icelandic texts, both in terms of grammar, but particularly in the model's understanding of Icelandic and its ability to translate Icelandic text to English.

Since the introduction of Chat GPT, the understanding of the model in Icelandic has been greatly enhanced by an active collaboration with the Icelandic language technology company Miðeind. The ability of Chat GPT to write quality Icelandic language texts has steadily improved, as have the possibilities to use it as a basis to develop further solutions in Icelandic. The improvements and progress in the model's Icelandic skills demonstrate how important it is to invest in training such models by developing training and test data.

Likewise, the collaboration with Open AI shows that much could be gained if other collaborations would be as successful, such as with Google, which recently introduced the language model Gemini.

A new Language Technology Programme needs to ensure significant investments in continuing to foster relationships with foreign tech companies, and make sure that Icelandic language technology solutions are known within these companies and that the companies understand how easy it is to make Icelandic one of the languages offered in their solutions. It must also be ensured that Icelandic is part of European cooperation projects in language technology, which can involve major grants from the EU. The strength of the Icelandic language technology industry and its value at the international level have already been proven. In November 2023, it was announced that the University of Iceland and Miðeind had received a major grant from Horizon Europe, the EU's funding programme for research and innovation, to take part in a project to develop a large AI model for Germanic languages, including Icelandic. Participation in such projects can be of great benefit to Icelandic language technology and the future of the language.

A new Language Technology Programme needs to ensure significant investments in continuing to foster relationships with foreign tech companies [...]



Our Icelandic, Everywhere

Chapter 2



2. Operation of the Language Technology Programme

The fruits of this labour are obvious from the results achieved during the project period of the previous Language Technology Programme. This next phase will have different focus points, and it is important that all language technology frameworks and operations reflect these changes.

It is important to establish continuity in Icelandic language activities. The collaboration of institutions, universities and industry during the project period of the previous Language Technology Programme, and their well-defined roles in furthering Icelandic language technology, proved to be successful. The collaboration was sealed with the involvement of the private non-profit institution Almennarómur in the project. Almennarómur oversaw contracting for development of LT infrastructure and solutions on behalf of the Ministry. A special forum was formed for the development of core projects, the so-called SÍM Consortium, which led the work on developing identified core projects within the project plan Language Technology for Icelandic 2018-2022.

An Expert council composed of three foreign experts was also established to advise Almennarómur on the technical implementation of core projects. Thus, the Expert council reviewed both the technical specifications of the cooperation agreement between Almennarómur and SÍM in the context of the Language Technology Programme, and SÍM's progress reports before any payments for projects were made. Almennarómur also submitted regular reports to the Ministry on the status and progress of the Language Technology Programme.

In all major aspects, this arrangement proved suitable to ensure quality and progress, as well as ensuring optimal use of all resources.

The fruits of this labour are obvious from the results achieved during the project period of the previous Language Technology Programme. This next phase will have different focus points, and it is important that all language technology frameworks and operations reflect these changes.

2.1 Artificial intelligence and language technology

There is great overlap in the fields of AI and LT, each field being limited without the other. For the sake of simplicity, AI could be referred to as a kind of umbrella term while the field of LT is mostly a subfield thereunder. Policy-making in these areas has so far been separate, although both language technology and artificial intelligence are always referred to in the resulting policies. The Icelandic government's current policy on AI from 2021, for example, specifically addresses LT and reiterates the importance of continued progress in the field.

The policy also points out that Iceland lacks continuity and links from research and studies on the one hand, and the implementation and use of AI technologies on the other. This is an interesting point and consistent with the conclusions on utilisation of LT discussed in this report.

It is also clear that the administration and framework for both LT and AI in Iceland is less developed than in many comparable countries. Most of our neighboring countries have established special AI organisations that play a key role in this process. Interestingly, all the institutions are also involved in language technology. One key point is that these institutions are jointly supported by companies, universities, and the government.

The idea that a comparable institution be established in Iceland is not new and is, for example, one of the proposals put forward in Iceland's AI policy. To quote the policy:



To this end, use is made of so-called applied basic research, which has the distinction of being governed by the needs of the economy. Such research plays a role that neither universities nor the industries can perform on their own: making practical use of technology and ideas and transferring solutions from one industry or market to another. The transfer of AI opportunities and technologies across industries, markets and sectors does not happen spontaneously, especially for SMEs with limited specialisation and capital. The majority of companies in Iceland are such companies.

This is in line with the key priorities on which the next Language Technology Programme will be based. It is helpful to look further into AI institutes in neighbouring countries to see the potential opportunities in a powerful AI and LT institution in Iceland.

2.2 An Icelandic LT and AI centre

In preparing this report, particular attention was paid to the arrangements for language technology and artificial intelligence in the Nordic countries. The group met with Alexandra Institute, a Danish private non-profit institution specialising in LT and AI research and practical solutions, AI Sweden, which has a similar role in Sweden, and the Norwegian entity NorwAI. A similar entity also operates in Finland.¹ The legal basis of these institutions differs, as does their funding.

Alexandra Institute is one of seven institutions in Denmark recognised as an RTO (research and technology organisation) as defined by the Danish Minister of Education and Science. All are private non-profit institutions with the role of researching the latest technologies and developing technology solutions that benefit Danish industries. Alexandra is the RTO in Denmark specialising in LT and AI development and consulting for companies.

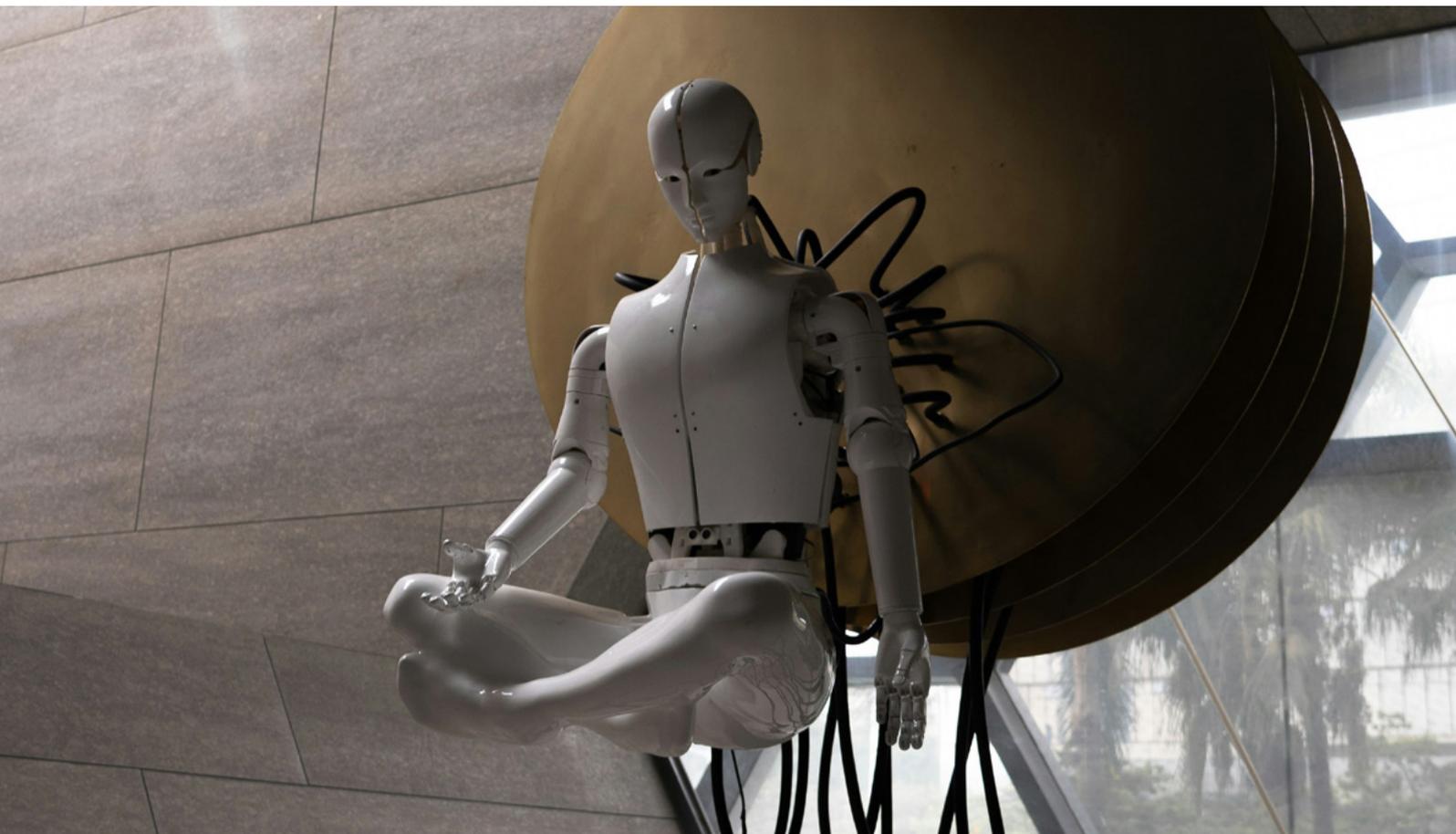
Approximately one-third of the Institute's funding comes from the Danish Ministry of Education and Science, one-third from EU grants for specific projects, and one-third from Danish industry through joint development of solutions, research and consultancy in the implementation of solutions. The Institute aims to increase the competitiveness of Danish industry by providing state-of-the-art LT and AI solutions in Danish, which for most companies would be impossible to develop on their own due to the complexity of research and the high costs involved. In this way, Danish companies have the chance to implement the solutions at a lower cost as they are largely developed with the involvement of the Danish government and the European Union. The services of RTOs are mostly used by small and medium-sized Danish enterprises. At a meeting of the group with Jens Kaas Benner, head of Alexandra's Copenhagen office, he explained that the main task of the Institute was to ensure the utilisation of LT and AI solutions for the Danish public and industries, which is also Iceland's biggest current challenge.

AI Sweden is a more research-oriented institution and has a different approach to its implementation role than Alexandra. Although AI Sweden is also a private non-profit institution, its regulatory environment is not comparable to that of the Danish RTOs. Half of the Institute's funding comes from the state-owned innovation institution Vinnova while the other half comes from approximately 120 Swedish companies and universities, big and small, that form the private non-profit institution AI Sweden.

The institution defines its role as producing practical LT and AI solutions, which will subsequently be put into use, rather than focusing its own efforts on the general implementation of such solutions, like Alexandra Institute does.

¹ The group was unable to meet with the Finnish Centre for Artificial Intelligence while preparing this report, but has initiated communications with the Centre that the group's contact, an employee of the Ministry of Culture and Business Affairs, could take further.

Both institutions bring together researchers and those who are ultimately going to make use of the solutions that the research aims to support. They bridge the gap between infrastructure and end-users; scientists and the economy. They both support the aim of developing linguistic solutions for smaller languages such as Danish and Swedish, funded by their governments, European project grants and contributions from industries. Such funding arrangements are necessary in linguistic areas that serve market areas as small as those in all the Nordic countries. The institutions then communicate closely with economic entities and do not embark upon projects or the development of solutions before engaging in extensive dialogue with potential users and analysing the needs for the solutions in question. All takes place under one roof; the institutions employ scientists and researchers to work on applied research and technological development, as well as providing the solutions to end users and collaborating closely with other researchers, such as universities, and industry. Furthermore, all solutions developed according to this arrangement are open-source and can therefore be used more widely in society. These institutions also bring together research and development in language technology on the one hand, and artificial intelligence on the other, which is a preferable approach going forward since AI and LT have become so closely entwined.



This arrangement is different from the arrangement in Iceland during the project period of the previous Language Technology Programme, where the government defined the tasks to be undertaken by the language technology community over a period of four years, and then entrusted the private non-profit institution Almannarómur with acting as intermediary for these projects and monitoring their execution. The technical development of these projects was then carried out outside Almannarómur, either at universities or language technology companies. This meant that the knowledge generated by the development is retained within these universities and companies as well as the actual solutions developed, although they are indeed published under open-source licences. Furthermore, Almannarómur has not had the defined role of ensuring the involvement of industries in projects or ensuring the development of solutions suitable for, and in close consultation with, relevant industries.

Although this arrangement proved successful during the project period of the previous Language Technology Programme, as has been noted, the steering group considers that the objectives of the next programme will be better served by other means.

In the group's opinion, the government should preferably aim to establish an AI and LT institution in Iceland, comparable to the ones found in the Nordic Countries and across Europe. Some of the main benefits would include:

-
- An established venue for AI and language technology in Iceland. This would ensure the responsibility, continuity and stability of both fields.
-
- Funding arrangements whereby funding is divided between the government, private companies and international grants would provide a great incentive for applying the technology.
-
- Sharing important resources, such as processing capacity, would be easier.
-
- Continuous development of knowledge in the field of AI and LT and dissemination of this knowledge within the institution. This is especially important given how few people work in the field in Iceland and this way, a kind of “institutional memory” could be established.
-

Many factors must obviously be examined further: what arrangements are most suitable in Iceland, whether the government should be directly involved in such an institution or support a private non-profit institution similar to the one in Denmark, and whether private companies are interested in providing funding for such an institution, as well as carrying out a thorough examination of the legal environment in Iceland for such institutions compared to the situation in the other Nordic countries. It is also clear that it will be a difficult and time-consuming task to establish such an institution, should that be the conclusion once all the analysis has been concluded. With this in mind, the steering group recommends that the Ministry of Culture and Business Affairs immediately seek cooperation with the Ministry of Higher Education, Science and Innovation, which is responsible for issues relating to artificial intelligence, to conduct a feasibility study on establishing an LT and AI institution in Iceland according to the Nordic models. The study will provide a comprehensive overview of the management of such institutions in the Nordic countries, potential increase in efficiency, the benefits of combining AI and LT, and whether such an institution could streamline and provide more flexibility in the development of LT solutions for end-users, and possibly ensure funding for projects from other sources than the Treasury. The feasibility study would be carried out as early as this year, 2024, and submitted to the ministries along with a list of proposed actions. The ministries would subsequently decide whether to establish such an institution and combine such efforts with preparations for the next Language Technology Programme.

2.3 Administration in the coming years

However, while such an LT and AI centre is not in operation, the operation of the proposals submitted in the report up to 2026 needs to be ensured, and the steering group recommends that these projects be outsourced to a single party or institution, which would then be responsible for the following:

- 1 Project management of the Language Technology Programme. The party in question will be entrusted with the role that Almannarómur held during the project period of the previous Programme, including managing the contracting for core projects on behalf of the Ministry.
- 2 Managing the allocation of utilisation grants.
- 3 Promotional and advisory roles on behalf of the government and work towards introducing Icelandic language technology to the Icelandic economy and in the solutions of foreign tech companies.

The scope of these projects is considerable, and the group proposes that the Ministry of Culture and Business Affairs provide at least ISK 85 million annually for these operations, which are also a highly significant part of the Programme's utilisation aspect. These aspects of the operations are discussed in more detail in subsequent chapters of the report.

Almannarómur has accumulated important experience and knowledge from the activities of the previous Language Technology Programme, and this arrangement has proved its worth. It is therefore recommended that Almannarómur become the centre for LT in Iceland throughout the term of the next Programme, and that Almannarómur be responsible for ensuring that the Programme's projects are carried out by the experts, institutions and companies that are most qualified to do so. This is addition to the constant work towards the Programme's principal aim: to support and ensure the utilisation of Icelandic language technology for the benefit of the public.

The project manager at Almannarómur will monitor the execution of core projects and maintenance of infrastructure. Almannarómur would also oversee the utilisation grants and be involved in their allocation in collaboration with the Ministry of Culture and Business Affairs, as well as handling the promotion and export of Icelandic language technology. Handling all these issues in the same place creates a good overview of the Programme's overall objectives. Services to companies and consultancy on the implementation of LT solutions should also increase greatly and provide direct links to the LT solution research community.

2.4 Funding arrangements for language technology

Although developing infrastructure through the core projects of the previous Language Technology Programme proved successful, it is now essential to implement comprehensive management of funds and projects. Funding for the previous Language Technology Programme was mostly provided through two basic channels: either by direct contributions from the Ministry of Culture and Business Affairs (previously the Ministry of Education, Science and Culture) or funding from competitive funds, i.e. the Strategic Research and Development Programme for Language Technology, which has been allocated via the Icelandic Centre for Research.

The Strategic Research and Development Programme for Language Technology is an open competitive research fund that operates according to the Act on Public Support for Scientific Research. Funding is awarded in accordance with the aims established by the fund's Board and based on an expert assessment of the quality of projects,

the reach of the project, the need for the deliverables of the project, the capability of the individuals carrying out the proposed research, and their ability to devote time and effort to the project. The purpose of the Strategic Programme is to facilitate the use of new communication technologies in Icelandic society, for the benefit of the public, institutions and companies.

During the five-year project period of the previous Language Technology Programme, 2018-2022, a total of ISK 1,713 million came directly from the Ministry of Culture and Business Affairs (formerly the Ministry of Education and Culture) and approximately ISK 542 million came through the Strategic Research and Development Programme for Language Technology. It was expected that the core projects of the Language Technology Programme would be funded by the Ministry, but that alongside this work, a special utilisation fund would be set up to support LT projects that were intended to make use of the infrastructure developed under the core project part of the Programme.

It is obvious that the utilisation fund proposed in the previous Language Technology Programme, which was to support the development of end solutions and the utilisation of LT infrastructure with a corresponding contribution from the business community, never materialised. The grants that have been awarded for language technology projects through the Strategic Research and Development Programme have not been successful in achieving the objectives of participation by the business community. Nearly all projects that received funding from the Strategic Research and Development Programme were, in whole or in part, carried out by entities in the research community working on developing LT infrastructure. The involvement of companies in the consumer market has thus been negligible.

It is important to better ensure the involvement of the business community in language technology, both in terms of the objectives of the Language Technology Programme and, not least, to encourage companies to invest in the utilisation of LT and AI.

One of the biggest challenges faced by Icelandic language technology is the efficient delivery of language technology products to users. The current arrangement of funding for projects does not fully meet the challenges that lie ahead for Icelandic language technology. It is important to review the arrangement of funding, both to give the government a better overview of the projects being carried out, and to ensure that funding is provided in accordance with the government's policy on the next Language Technology Programme's priority projects. In view of this, it is prudent that funds for utilisation projects should be provided to a greater extent in a centralised manner rather than for the Strategic Research and Development Programme to play that role. In doing so, projects could be better prioritised, in relation to the strategy to be developed, ensuring that sufficient resources will be spent to increase the use of the LT products developed through the core project parts of the Programme. Much more attention needs to be paid to the utilisation of the infrastructure that has already been developed, and therefore the funding framework needs to take this into account.

Thus, it is proposed that the Strategic Research and Development Programme for Language Technology will remain unchanged through 2026 until a complete overview of Iceland's LT and AI landscape has been obtained. In the meantime, special utilisation grants for Icelandic language technology will be established, looking both at grants for end solution projects that use LT infrastructures, and at support for the general implementation of these infrastructures into existing systems.



Chapter 3



3. Utilisation of language technology

Although the development of infrastructure for Icelandic language technology and collaboration between universities, institutions, companies and non-governmental organisations, under the aegis of Almannarómur, have been successful, the future of Icelandic in a digital world is still far from guaranteed. Without continued targeted action, its future remains uncertain. It is therefore very important to further develop the language technology infrastructure in which investments have already been made.

The main focus of the previous Language Technology Programme was to develop Icelandic LT infrastructure to ensure that Icelandic would be eligible in a digital world. Much has been accomplished during this time: An Icelandic speech recogniser, synthesiser, basic machine translation between Icelandic and English, revision tools and much more was developed as part of the previous Programme. Now is the time to ensure that this infrastructure is put into widespread use, both among the general public in everyday life, and in business and government.

In this, three factors are most important:

- 1 That Icelandic organisations wishing to implement or develop new end solutions, based on open-source LT infrastructure, can apply for grants to complement their own contributions.
- 2 That Icelandic organisations wishing to implement or deploy open-source LT infrastructure can receive advice and support for targeted implementation.
- 3 Continued communication with foreign technology companies that develop solutions that are, or may become, widespread among the Icelandic public, and work systematically to bring Icelandic language technology to the solutions of these companies.

3.1 Funding for utilisation projects

As mentioned in Chapter 2.4 above, the utilisation fund proposed in the previous Language Technology Programme was never realised, so necessary support for utilisation projects has therefore been lacking. For such grants, Icelandic companies could apply for funding to implement LT solutions in their services or use LT infrastructure to further develop practical end solutions for users. Such grants have been sorely lacking in the current grant environment in Iceland, and language technology companies, education technology companies and others who have expressed an interest in utilising Icelandic language technology have complained that they have not been able to apply for the necessary grants for their projects.

It is recommended that the Ministry of Culture and Business Affairs contribute ISK 60 million annually to utilisation projects. Almannarómur would manage funding, request applications and disburse the grants that are allocated. An allocation committee will be appointed to assess applications and the projects to be supported, composed of the project manager for the Language Technology Programme, a representative from Almannarómur and a representative appointed by the Minister of Culture and Business Affairs.

It will be ensured that not only full-fledged development projects that deliver end solutions for users will be supported, but also that applications can be made for the implementation of existing LT solutions in the systems and services of businesses.

Eligibility for grants would be restricted to the applied use of LT infrastructure and products in projects aimed at enabling users of digital solutions to use Icelandic. Grants would require a 50% reciprocal contribution, meaning that Icelandic business would thus take an active part in supporting the Icelandic language technology industry.

It is suggested that to begin with, particular attention is paid to support solutions that can be applied within the health or education system. It is quite clear how language technology can be utilised within these sectors, and solutions have already emerged that make use of LT developed specifically for these sectors. Efforts will also be made to strengthen solutions that can be used to teach Icelandic as a second language, both within the school system and for people in the labour market, making use of Icelandic LT to do so. When awarding grants, the Language Technology Programme's primary objective will be to ensure that the future of the Icelandic language in a digital world will always be the main consideration, and that the projects are likely to strengthen the position of the Icelandic language in society.

3.2 Counselling for Icelandic industries and general introduction of language technology solutions

In the course of the consultation and analysis conducted by the steering group during the initial weeks of its appointment, it soon became clear that the work carried out under the previous Language Technology Programme could have been better promoted. Only one employee of Almannarómur, which oversaw the operation of the Programme, worked on the Programme during the project period, and it was unclear who was responsible for promoting the LT solutions developed under the Programme; the companies that developed solutions, Almannarómur or the Ministry of Culture and Business Affairs.

As a result, Icelandic business, non-governmental organisations and the public could have been better informed of the possibilities offered by Icelandic LT infrastructure.

Nor was there any central consultancy for companies interested in continuing to develop end solutions. As a result, there was limited information and advice on which infrastructure versions suited them best, which infrastructure updates were the latest and which Icelandic language technology companies were best suited to assist them in the development of the end solution.

It is suggested that Almannarómur take on this role and that one of its employees assume the role of advising Icelandic companies and organisations and assisting in the implementation of LT solutions. Almannarómur would also promote Icelandic LT to the public and the business sector, oversee conferences, etc.

It is furthermore proposed that Almannarómur receive funding for updating its website to better present what is currently available in Icelandic LT and what the policy is for the next few years.

3.3. Communication with foreign tech companies

In recent years, efforts by Almannarómur, the Icelandic government and Icelandic language technology companies to introduce Icelandic language technology to solutions from foreign tech companies have yielded some results.

Continued work on this project is obviously one of the most important factors in ensuring the use of Icelandic language technology by the general public, thereby ensuring the digital future of the language in a world of extremely rapid technological development.

Continued work on this project is obviously one of the most important factors in ensuring the use of Icelandic language technology by the general public, thereby ensuring the digital future of the language in a world of extremely rapid technological development. There is no way to ensure that companies will adopt Icelandic language technology,

but experience in recent years has shown that most companies are open to it if it does not entail too much effort or expense. It is therefore important that targeted efforts be made to meet with foreign tech companies to discuss Icelandic language technology solutions and make sure that Icelandic is included in their solutions.

OpenAI

The collaboration between the Icelandic language technology company Miðeind and the US tech company OpenAI is an indication of the quality of the LT infrastructure developed in Iceland in recent years under the previous Language Technology Programme, and how this infrastructure can be utilised in a focused manner. The collaboration began with a May 2022 meeting in San Francisco between the Icelandic Presidential Delegation, the Minister of Culture and Business Affairs and Almannarómur with Sam Altman, CEO and founder of OpenAI. During this trip, Icelandic language technology was introduced to foreign technology companies with the aim of integrating Icelandic into as many everyday technology solutions as possible. Following the delegation's meeting with OpenAI, the company decided to choose Icelandic as the first language outside of English in which the AI model Chat GPT-4 was specifically trained. This was made possible with the help of Icelandic LT solutions. The model was launched in mid-March 2023 and attracted worldwide attention due to its conversational and question-answering skills.

Microsoft

In the spring of 2022, Microsoft released the voice synthesiser voices Guðrún and Gunnar. The voices are both operating system voices and accessible through Microsoft cloud services, so Microsoft's Edge browser can, like Google's Chrome, recite Icelandic text.

At the same time, Microsoft released its first Icelandic speech recogniser, enabling users to talk to the Microsoft operating system in Icelandic. One of Microsoft's team leaders in the project, Jacky Kang, confirmed during a meeting with Reykjavik University that open-source data from the Language Technology for Icelandic 2018-2022 project plan were used in the development of the speech recogniser. Reykjavik University has also provided suggestions to Microsoft on how to improve the quality of Guðrún's and Gunnar's voices.

Google

Google has released an Icelandic voice synthesiser on three occasions. It is likely that pressure from Reykjavik University and the Icelandic Association of the Visually

Impaired, as well as the relationship between Icelandic language technologists and an Icelandic employee at Google, resulted in the availability of an Icelandic voice synthesiser in Google's solutions.

In late summer 2012, Google introduced a new speech recogniser for Icelandic, which has been used for Android mobile devices, among other things. This technology converts speech into written text and can be used for voice operation of various Android devices.

In the autumn of 2015, Google released the voice synthesiser Anna, based on voice recordings made by Reykjavik University in collaboration with Google. This development therefore took place before the launch of the previous Language Technology Programme. Google has not provided information on whether Icelandic LT solutions released under the previous Language Technology Programme have been used for updates and further development of the speech synthesiser.

Experience has shown that much can be gained with good relations with foreign tech giants and initiatives in presenting them with Icelandic LT solutions that they can obtain and implement in their systems. It is therefore highly important that LT infrastructure developed under the next Programme will continue to be released under open licences and that the responsibility for continuing collaboration and communication with foreign tech companies be placed with one entity that keeps a close eye on international developments and takes the initiative to bring Icelandic LT to the table. It is also important that test data be developed for existing language technology infrastructure, as it is clear that the companies must be able to understand the quality of the solutions they are adopting, prerequisite of which is the availability of test data.

To ensure the future of Icelandic in a digital world, it is essential to integrate Icelandic into the solutions of companies such as Microsoft, Google, Meta, OpenAI and numerous other foreign tech giants. Many lesser-known companies are also offering web services that are widely used in business and education. These must also be mapped out to make sure that Icelandic is a part of the solutions offered.

It is suggested that Almennarómur be responsible for this aspect and that considerable efforts should be allocated for it in any service agreement between Almennarómur and the Ministry. It is suggested that Almennarómur organise further visits of delegations to foreign tech companies to advocate for the introduction of Icelandic LT in their solutions. Almennarómur will also have the role of monitoring projects in development by these tech giants and assessing whether and when a response is needed to ensure the status of Icelandic in these projects.

Chapter 4



4. Core Projects

4.1 Programme flexibility and new arrangements

During the project period of the previous Language Technology Programme, various infrastructure and solutions were made available under open-source licences. It is difficult to measure the quality of these solutions as solid test data have not yet been developed and published. It is clear, however, that the infrastructure can still be improved with continued development under the next Language Technology Programme.

With the rapid development of technology and the recent leaps in the capabilities of large language models, the environment has changed dramatically. The need for core solutions within language technology is different now than it was during the project period of the previous Language Technology Programme. There has been both a need for new core projects that were not foreseen in the previous Programme and varying degrees of necessary improvement to solutions that have been developed.

It is therefore essential that funding arrangements for the continued development of LT infrastructure are flexible and that rather than permanently linking funding to long-term projects, it will be possible to respond to changing technological circumstances by funding relevant new and urgent projects.

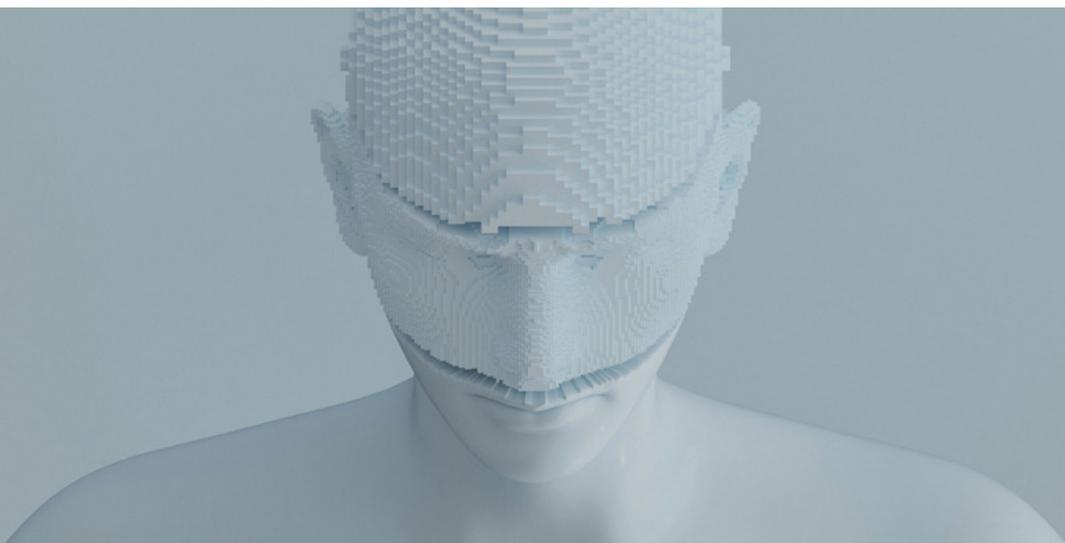
The uncertainties associated with rapid technological change must also be taken into account. The importance and necessity of certain components in the development of core solutions can become obsolete overnight with the advent of new and unforeseen external AI solutions. It is therefore essential that funding arrangements for the continued development of LT infrastructure are flexible and that rather than permanently linking funding to long-term projects, it will be possible to respond to changing technological circumstances by funding relevant new and urgent projects.

The steering group considers that in order to ensure flexibility in the development of LT infrastructure in coming years, it is necessary to change the funding arrangements for core projects.

The group proposes that contributions to core projects amount to at least ISK 100 million per year, and that the funding requirements of core projects be reevaluated annually by the Ministry of Culture and Business Affairs in cooperation with Almanna­rómur. Almanna­rómur will be contracted to ensure the execution of the projects by outsourcing them through contracts to SÍM for their execution, as was done during the project period of the previous Programme. It is assumed that contributions to core projects will continue to be exempt from VAT, and that all products will continue to be released under open-source software licences at no cost to users. Almanna­rómur will continue its role in monitoring the progress of the projects and regularly informing the Ministry of their status. The Expert council of foreign experts will continue their role and review contracts and progress reports on the projects. In addition, a special project manager for the Language Technology Programme, employed at Almanna­rómur, will be appointed.

It is proposed that new arrangements be adopted for the selection of each year's core projects to further increase the flexibility of the programme.

A project management board for core projects and their maintenance will be established, composed of two representatives of Almanna­rómur, two representatives of SÍM and one from the Ministry of Culture and Business Affairs. The project management board will convene once the Ministry of Culture and Business Affairs has finalised the amount for funding of core projects in the coming year, in order to assess whether the projects identified in the new project plan are still eligible.



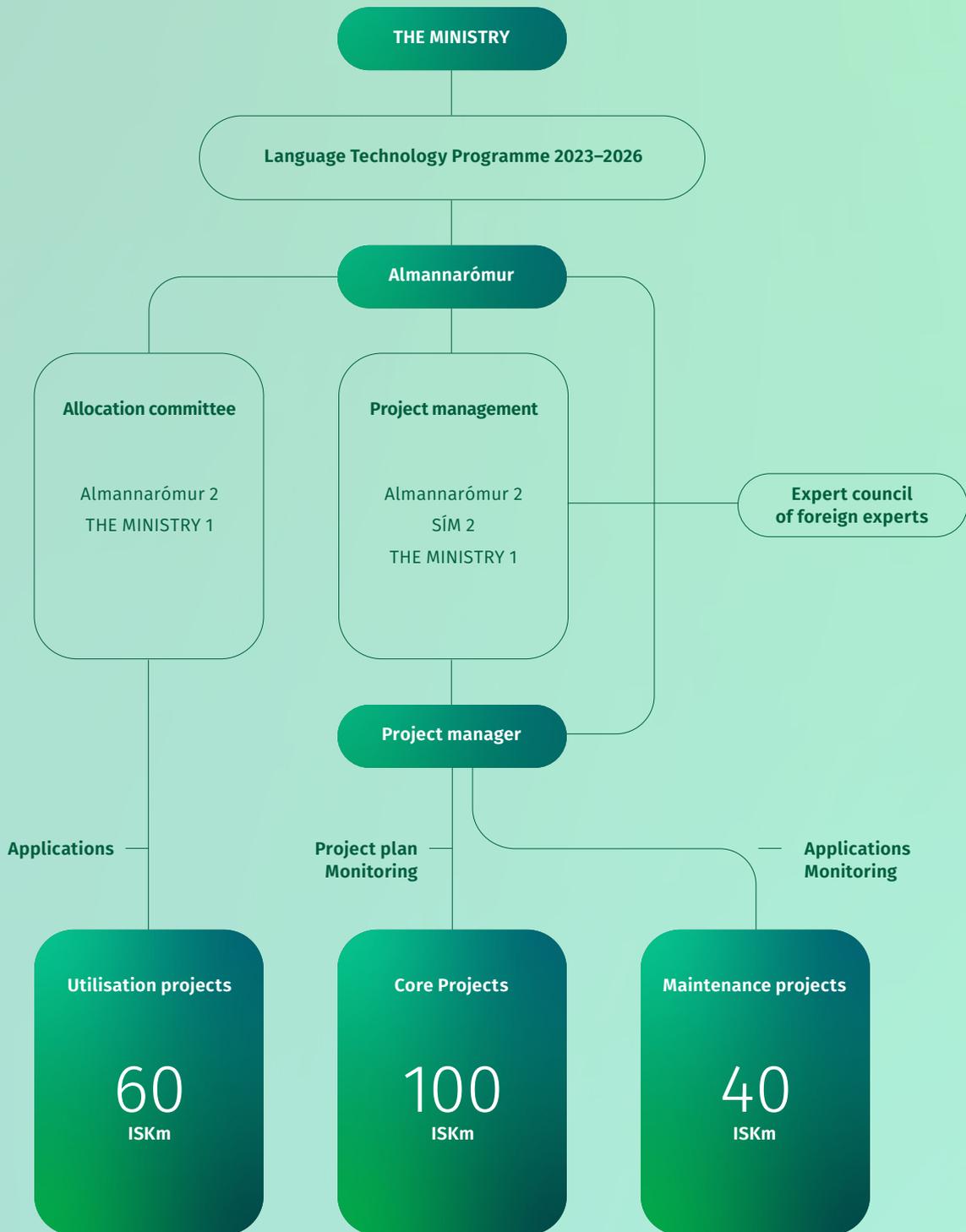


Figure 1: The figure shows the administrative organisation of the Language Technology Programme. Almannarómur is responsible for the administration of major projects. The project manager employed at Almannarómur oversees core and maintenance projects and consults with the project management board on decisions.

4.2 Core projects and competition considerations

It must be ensured that the development of core projects with funding from the government's Language Technology Programme does not impair the competitive standing of language technology companies operating in the Icelandic market. A clear distinction must be made between what is to be classified as a core project, issued under open-source licenses and entirely financed under the Language Technology Programme, and projects that are more suited to a competitive market. Core projects should thus be defined as the development of infrastructure that can be used in the development of end solutions and equally benefits all those who wish to undertake such work.

The implementation of LT infrastructure or the applied use of LT solutions will be financed from the utilisation part of the Programme, subject to a corresponding contribution from companies. The development of LT infrastructures has come a long way since the previous Language Technology Programme, and many of these are already ready for implementation in end solutions. It is important to focus the future development of core projects on infrastructure that can be successfully applied directly to end solutions developed in a competitive market.

4.3 Prioritised core projects

The steering group recommends putting the following core projects at the forefront of the new project plan, and that a team of experts be assigned to re-evaluate and prioritise them during project period. It is important to prioritise and consider funding some projects in other ways, including through European cooperation.

Language resources

Purpose: Language resources are the data on which language technology tools are based. To develop the tools and adapt the technology to Icelandic, varied and extensive data are needed. They are used to train the tools, but also to test them and assess their quality and accuracy. Quality assessment is important at all stages, both in development to ensure that the correct course is taken, and when the final product is ready. In this way, the user can be informed about the quality of the tools, whether it is the end user or someone using the tools in the development of other software. Some datasets, such as the Icelandic Gigaword Corpus, are useful for various projects in other components of language technology development while some projects require more specialised datasets. The project components are the following:

-
- Expanding the Icelandic Gigaword Corpus
-
- Parallel sentences in two or more languages
-
- The Icelandic Parsed Historical Corpus (IcePaHC) converted to the UD (Universal Dependencies) scheme
-
- Corpus for sentiment analysis
-
- Corpus for information extraction
-
- Corpus for summarisation with extract and from a longer context
-
- Corpus for question answering
-
- Corpus for word-sense disambiguation
-
- Corpus for paraphrasing
-
- Corpus for analysing difficulty level and simplifying text
-
- Multi-level corpus for sign language
-
- Multi-level corpus for in-person conversations (videos of conversations)
-
- Multi-level corpus with pictures and text descriptions
-
- Test data for coreference and anaphora resolution
-
- Test data for capability of large language models (language comprehension)
-
- Control data for bias analysis (e.g. gender bias)
-

Text processing and communications

Purpose: The analysis of texts with respect to various semantic and linguistic aspects is the basis for adapting language technology to the needs of specific groups or domains. Ambiguities need to be resolved to facilitate semantic analysis and the retrieval of information from the text. It is also necessary to facilitate terminology analysis to ensure the correct use of terminology in specialised texts. When text is directed at a specific group, it may be necessary to analyse its difficulty level in order to bring it to the appropriate register and make its content more accessible, whether for children, non-native speakers or other groups. The project components are the following:

-
- Coreference and anaphora resolution
-
- Automatic term extraction
-
- Automatic dictionary creation
-
- Text difficulty and text simplification
-
- LT solutions for teaching Icelandic and supporting non-native speakers
-
- Developing Universal Dependencies parsers
-

Speech recognition

Purpose: To extend the use of speech recognition to as many social groups as possible, such as children and non-native speakers, through continued development, and to continue developing speech recognition for various software and hardware environments, such as operating systems and smartphones. These project components are:

-
- Speech recognition for children, non-native speakers, and older voices
-
- Write mode for Icelandic speech recognition, such as commands for punctuation and line spacing
-
- Detection of code-switching
-
- Data collection in Icelandic via LibriVox (an open content, global audio library that provides access to readings of longer texts rather than individual sentences/shorter text segments)
-
- Speaker recognition for Icelandic
-
- Speech recognition for natural conversation, including back channel vocalisation
-
- Language and acoustic modelling for Icelandic
-

Speech synthesis

Purpose: For Icelandic speech synthesisers to be usable for those who need them in their daily lives (accessibility, security and rights issues); for Icelandic speech synthesisers to be usable in business and service environments; for Icelandic speech synthesisers to be usable for reading textbooks and other practical texts and publish them as audiobooks, thus facilitating the publication of content from entities such as the Icelandic Audiobook Library and the Directorate of Education. The project components are the following:

-
- Continued development and updates of speech synthesisers and vocoders
-
- Improving the voice synthesis for Android devices, which the blind and visually impaired use as an aid for daily activities
-
- Speech synthesis with child voices
-
- Voice adjustment and mixing
-
- Improved reading of text containing mixed languages and specialised terminology
-
- Specialised text normalisation
-
- Intonation and cadence
-
- Emphasis and intonation control in recited text
-
- User interface: additions/improvements to dictionary, abbreviations added, etc.
-
- Speech synthesis of longer texts/books
-

Machine translation

Purpose: Machine translation (MT) is important in order to make Icelandic texts available to non-speakers and to give Icelandic speakers the opportunity to understand foreign-language texts, e.g. in the media and online. Thus, MT is important for balancing and increasing the public's access to information. MT is also useful for specialised professional translations, for example translations of regulations at the Ministry for Foreign Affairs Translation Centre and for public and private service websites. MT will be useful for automatic captioning of foreign-language television material. An Icelandic translation engine, where linguistic quality and nuances in Icelandic are considered, including with reference to biases and hate speech, is an essential requirement. The project components are the following:

-
- Translations of longer-context and more diverse texts than is currently the case (more domains than news and regulations)
-
- Alignment of longer documents for Icelandic and languages other than English
-
- MT biases and their assessment
-
- Real-time translation of speech and conversation
-
- Automatic captions in Icelandic for foreign-language television material
-
- Use of specialised dictionaries/translation memories (TMs) in MT
-
- Identification and use of idioms in MT
-
- Lighter translation models to be used for smartphones without an internet connection, etc.
-
- Cross-language information retrieval
-
- Collaboration with WMT on Icelandic in shared tasks
-

Spell and grammar checking

Purpose: The aim of the project is for the Icelandic public to have easy access to a tool that corrects both the spelling and grammar of texts according to the most common language standards and norms. Such a tool must be accurate and able to serve different user groups, including non-native speakers, people with dyslexia and children. It is important that the tool is also available in different formats, including in the most common word processing systems and smart devices. Spell and grammar checking can be used to correct differences in advantage in society when it comes to fluency in Icelandic and associated prejudices. In business, spell and grammar checking can be useful in maintaining a company's tone of voice and ensuring that the language of communications to customers and other stakeholders is clear and consistent with linguistic standards. The project components are the following:

-
- Continuing development of spell and grammar checking with annotated suggestions
-
- Continued development of spell and grammar checking with rapid conversion of text with errors to better text, where explanations of corrections are not required
-
- Additions to the language standard (additional items covered, detailed instructions for each item, etc.)
-
- Improving the general error corpus
-
- Dyslexia error corpus
-
- Child language error corpus
-
- Non-native speaker error corpus
-
- Custom spell and grammar checking for user groups, such as non-native speakers, people with dyslexia and children
-
- Identification and correction of more complex issues (gendered language, language register and style, word choice)
-
- Continued development of additions; integration of spell and grammar checking into software, operating systems and devices
-

Artificial intelligence (AI) and dialogue systems

Purpose: Utilising the latest AI technologies, including large language models, to develop various resources for Icelandic, comparable to those available for larger languages. These include open questioning answering, text summarisation, text simplification, sentiment analysis, contact centre assistance, etc. Such resources can also combine speech recognition, speech synthesis and speaker diarisation to create comprehensive dialogue systems. Objectives include being able to speak to the technology in Icelandic and for the technology to comprehend and respond as naturally as possible. At the same time, technology can enable businesses to deploy AI in various internal processes based on textual and statistical data and to assist managers.

The adaptation of AI and dialogue solutions to Icelandic is a key factor in defending the position of the Icelandic language in the fourth industrial revolution, and in ensuring that future devices and software can understand, process and convey information in Icelandic instead of English taking over.

-
- Large language models adapted for Icelandic and used for various resources, including:
 - » Open question answering and information retrieval
 - » Text summarisation, text simplification and information extraction
 - » Sentiment analysis
 - » Chatbots/contact centres
 - Dialogue systems, e.g. Embla/Siri/Alexa/Cortana, partly based on AI
 - » Dialogue manager
 - » Natural language understanding
 - » Natural language generation
 - » Various parts of dialogue not included in the text, such as filler words and cadence
 - Minimising biases and errors in data and models; measurements; reports
-

4.4 Maintenance of infrastructure and core solutions

The Language Technology Programme includes the hosting and maintenance of data and LT infrastructure developed during the project period of the previous Language Technology Programme, as well as those to be developed under the next Programme. An agreement has been made with the Árni Magnússon Institute for Icelandic Studies for the operation of a CLARIN centre in Iceland to the end of 2026.

CLARIN ERIC is a European forum for cooperation on research infrastructure in language technology. CLARIN stands for *Common Language Resources and Technology Infrastructure* and **ERIC** for *European Research Infrastructure Consortium* which is the legal form of the cooperation network. The objective of the CLARIN network is to support the development, recording, maintenance, retention, use and sharing of language data and resources for research of such data in the humanities and social sciences.

It is recommended that the arrangements for the operation of an Icelandic CLARIN Centre within the Árni Magnússon Institute should be reviewed before the end of this Language Technology Programme in 2026, and that it would be desirable to have all Icelandic LT projects in one central location, as detailed in Chapter 2.2.

Care must be taken in the maintenance of existing LT infrastructure so that it does not become obsolete with rapid technological developments or updated programming languages or operating systems. The group recommends abandoning the arrangement of allocating predetermined funding for the maintenance of individual projects. Instead, the projects listed below will be merged into one website, hosted by Almennarómur, and maintenance and updates will be carried out at the discretion of the Programme's project manager in consultation with the project management board, as described in Chapter 4.1.

The project management board would have an annual budget of ISK 40 million to allocate as needed and according to the prioritisation of solutions in need of maintenance. Thus, the project management board would allocate funding to institutions and language technology companies operating LT infrastructure. The project manager will work closely with SÍM and other Language Technology Programme entities to assess the maintenance needs of solutions and forward applications for maintenance grants to the project management board, which will then decide which core solutions will receive funding for maintenance.

Maintenance tasks will be prioritised based on factors including their number of users, their scope, and the urgency of updates and amendments at each time. Maintenance funds are allocated on the condition that funding be used for the maintenance of data or infrastructure that has been developed with funds under the first or second Language Technology Programme and issued under open-source licences.

This would allow the focus on maintenance to be prioritised so that the most used projects will have priority instead of funding being tied in advance in potentially less used projects or projects that may become obsolete with time.

It is also considered preferable to maintain particular web services that have been established to provide access to core project products, and host and operate such services in a single location.

