

Kone
Foundation

*The Language
Programme
2012–2016*



Kone Foundation
The Language Programme

KONEEN SÄÄTIÖ

I.2.2012

GRAPHIC DESIGN
Patrik Söderlund

PRINTER
Mynäprint, Mynämäki 2012

The Kone Foundation Language Programme 2012–2016

CONTENTS

Background	5
Kone Foundation Places Emphasis on the Minority Languages of the Baltic Region and on Finnish	5
Supporting Multilingualism with the Help of the Language Programme	6
The Current Support for Finno-Ugrian Minorities and Languages	7
Mission	12
Vision	12
The Mission: Documentation of Small Finno-Ugrian Languages, Finnish, and Minority Languages in Finland	14
Enhancing the Availability, Accessibility and Usability of Corpora	15
■ Taking into Consideration All Kinds of Language Users	17
Advancing a Culture of Openness and Interaction	17
Advancing Multidisciplinarity and Crossing New Boundaries	19
How Are the Objectives Achieved?	20
Concrete Measures 2012–2016	22
Communication and Cooperation between Projects	22
Projects Funded Through Annual Grant Application Rounds, Specially Allocated Grant Calls, and Projects Initiated by the Foundation	22
■ An Example of the Use and Development of Language Technology Applications	27
Monitoring the Programme	30



Background

KONE FOUNDATION PLACES EMPHASIS ON THE MINORITY LANGUAGES OF THE BALTIC REGION AND ON FINNISH

One of the key goals of the Kone Foundation is to strengthen the status of the Finnish language and of the minority languages in Finland and in the Baltic Region. Finnish, in particular, is taken into consideration in all the activities of the Foundation. The measures taken to strengthen the status of the Finnish language include not only the general support of Finnish non-fiction writing but an award dedicated to rewarding excellence in the field (*Vuoden Tiedekynä*, “science pen of the year”). Likewise, the interests of the minority languages of the region are taken into account for example in the allocation of grants to arts and culture projects. Furthermore, the foundation wishes that authors and translators writing in small minority languages would apply for residencies at Saari Manor.

However, the Kone Foundation now wishes to promote the status of Finno-Ugrian languages even more efficiently through a language programme dedicated to the task. Unfortunately, the amount of speakers of many such languages has decreased and several are endangered. Reasons behind these phenomena include poor basic linguistic rights and the dispersal of speech communities as a result of urbanization.

Another reason behind the weakened status of many languages is globalization. In fact, the ever decreasing use of Finnish as a scholarly language is an example of just this. Guaranteeing and developing thought requires the ability to discuss scientific issues also in Finnish. It is necessary to ensure that the basic concepts and terminology of different scientific fields also exist in one’s mother tongue. Despite the diminished status of the Finnish language in science, it is still a highly practical and functional language in all aspects of life. Unfortunately, this is not the case for most small Finno-Ugrian languages.

The status of the Finnish language in Finland is strong even within the context of the whole world, and fortunately, the government takes care to protect it. However, the impact of the financial crisis on the public sector has also held ramifications for the status of the Finnish language.

For example, the budget cuts included in the government's productivity programme affect the Institute for the Languages of Finland (Kotus). Furthermore, the financial problems of the public sector are also reflected in the status of Finnish in various government agencies and universities, as well as in the status of Finnish language classes in the revised curriculum.

Language is not only an important tool for conceptualizing the world around us but it is also an essential part, and reinforcer, of one's identity. From the point-of-view of the Kone Foundation, languages and the linguistic diversity of the world hold intrinsic value that should be protected and supported. Moreover, linguistic research, documentation and support are essential in understanding society, culture and the world, and such information can be profited from in many different ways.

In 2011, the Kone Foundation hired Jack Rueter, Ph.D., for six months to prepare the Language Programme and to gather relevant information.

SUPPORTING MULTILINGUALISM WITH THE HELP OF THE LANGUAGE PROGRAMME

Multilingualism refers to the ability to speak several languages at a native level. It is also sometimes understood in the broader sense of being able to communicate fluently in several languages. The active use of more than one language is more a norm than an exception. Almost 7,000 languages are spoken by the seven billion people inhabiting the Earth. The Baltic Sea Region, which includes Finland, is also a multilingual area. The variety of languages spoken in the region include not only Finno-Ugrian, Germanic and Slavic languages but also others that originate from further away or have developed only recently. In fact, most Swedish-speaking Finns also speak Finnish at a native level, and immigrants are another large multilingual group in modern Finland.

Some languages are legally recognized as official or dominant in their own countries whereas others lack official status. Minority languages are often at a disadvantage as proficiency in a dominant language is often a prerequisite for procuring a place of study or a job, for example. Young people moving to cities are in particular danger of abandoning their mother tongue if there are no opportunities for its use in the new environment.

Even today it is common for people to have misconceptions regarding multilingualism. Parents may, for example, stop speaking their mother tongue to their children for fear of bilingualism impeding their language development. Such misunderstandings lead to a decrease in minority language speakers and to the weakened status of languages.

Minority languages are a vital resource and represent the world view of the communities that speak them. Furthermore, the unique culture, traditions and world view of a minority are best communicated in the free expression of its native language. The aim of the Language Programme is, therefore, to promote the understanding that the use of several languages in life is not any kind of impediment to an individual. On the contrary, multilingualism is an asset. When people are able to speak many languages, they are also introduced to the consciousness that there is a multitude of ways to interpret the world, and such a consciousness can enrich the whole of society.

CURRENT SUPPORT FOR FINNO-UGRIC MINORITIES AND LANGUAGES

Many public and third sector organizations in Finland support small Finno-Ugrian minorities and their languages. The Foundation intends to find its own niche amongst these different organizations and to avoid overlaps. The idea is also to work together with others in the same field and to support them when necessary.

Finno-Ugrian minorities and languages are being studied at universities in Helsinki, Turku, Oulu and Eastern Finland (Joensuu). The following organisations also support and/or study peoples and languages to a significant degree:

»» **The Institute for the Languages of Finland, Kotus**, studies and maintains Finnish, Swedish, Sámi, Finnish Sign Language and Romani. However, Kotus' participation in the study and lexicography of Finno-Ugrian languages is minimal nowadays. Most of Kotus' research results are published as dictionaries. It is disconcerting that due to a lack of resources, the compilation of the Dictionary of Old Literary Finnish

was terminated and funding for the Dictionary of Finnish Dialects has been reduced. A dictionary project focusing on immigrant languages is, however, an example of a new project by Kotus.

»» **The Council for Finnish Studies at Universities abroad (UKAN)** sends language assistants to the universities of kindred Finno-Ugrian peoples to assist in the teaching of Finnish, as well as Finnish university teachers to lecture for short periods. In addition to Finnish language, the teaching also includes familiarization with Finnish culture. UKAN also organises literature translation seminars.

»» **The M.A. Castrén Society** was founded to create and promote contacts between Finland and the Uralic-speaking peoples of Russia, to spread knowledge of their unique cultures in Finland, and to support their cultural enterprises and the cultivation of their native languages. The Society also supports translations of literature.

»» Since 1994, the Society has been responsible for carrying out the **Kindred Peoples Programme** funded by the Finnish Ministry of Education. The programme primarily targets cultural, educational and research exchange as well as museum and library cooperation. The Kindred Peoples Programme funds research visits to Finland in the form of CIMO scholarships, for example. The programme will be renewed in the near future.

»» In 2007–2009, the **Finnish Cultural Foundation** supported the preservation of the languages and culture of kindred language speakers living in Russia with over 400,000 Euros. The projects included organizing Veps and Udmurt language camps, supporting museums and organizing multi-professional publication training. This large financing project began in the fall of 2006 and received additional funding of over a million Euros in 2008. The projects supported language nests, NGO training and the revival of the village of Haikola. Approximately twice a year, the Cultural Foundation calls together the participants of its projects and colleagues from within the field. The Kone Foundation has also participated in these events since the spring of 2011.

»» **The Finland-Russia Society** develops Finnish-Russian culture and citizen contacts, and participates in Finno-Ugrian operations by, for

example, assuming responsibility for the preparations for the World Congress of Finno-Ugrian Peoples in Finland. The Society has organised NGO training and language nests with funding from the Finnish Cultural Foundation and the Ministry of Foreign Affairs, as well as a Finno-Ugrian cultural cooperation project for youths in the Republic of Karelia with funding from the Unit for Regional Cooperation at the Ministry of Foreign Affairs.

»» **The Finno-Ugrian Society** promotes Uralic and Altaic languages as well as the culture, ethnology, national poetry and prehistoric research of the peoples that speak them with the help of grants and publication activities.

»» **Sukukansojen ystävät ry** (“Friends of kindred peoples”) spreads knowledge about Finno-Ugrian peoples in Finland, organises events and creates contacts between Finnish and kindred peoples’ NGOs and grassroots actors. The association is also the Finnish representative of the Youth Association of Finno-Ugrian Peoples, MAFUN.

»» **The Committee of Kindred Peoples at the Student Union of the University of Helsinki** also supports the aspirations of Finno-Ugrian peoples to nurture their language and culture. The Committee directs collected funds towards Finno-Ugrian minority projects. Fundraising proceeds have been used to support several dozens of projects planned and executed by active youths. Aid has been granted particularly for publication activities, equipment purchases, and the organisation of social and cultural events.

»» **The Finnish Sámi Parliament** is the self-governing body of the Sámi people. The Sámi Parliament Language Council upholds and develops Sámi languages used in Finland, gives instructions and recommendations on matters concerning grammar, nomenclature and terminology, promotes linguistic research, offers information on the Sámi language, and participates in Nordic language cooperation. The Sámi Parliament also offers financial support to Sámi organisations and Sámi arts and culture, and is responsible for compiling study material.

»» **The Academy of Finland** is planning a joint initiative on Finno-Ugrian languages and cultures with the Hungarian Academy of Sciences.

Furthermore, projects involved with the digitalisation of language material and/or language technology applications as well as other related projects are also important in terms of the Kone Foundation Language Programme. Examples of the above include:

»» **The Finnish Terminology Bank of Arts and Sciences, BFT**, is a project carried out at the University of Helsinki and funded by the Academy of Finland. This joint project between the University of Helsinki, Kotus, the Federation of Finnish Learned Societies, and The Finnish Terminology Centre TSK aims to create a foundation for a shared, freely accessible, multilingual terminology bank for the arts and sciences of Finland.

»» **The National Digital Library, KDK**, is a project initiated by the Ministry of Education that aims to promote the availability and usability of the central national information reserves of libraries, archives and museums, as well as to develop long-term storage solutions for cultural heritage material that is currently in digital form or in the information network.

»» *Kirjallisuuspankki* is a plan to establish a digital online resource of copyright-free fiction from Finland, in Finnish and in Swedish, following international models. Texts that have already been digitalised and proofread serve as a foundation for the service. The resource can be used for both research and language teaching (UKAN Finnish teachers abroad, for example).

»» **FIN-CLARIN** is the Finnish partner of the **ESFRI** project, which is developing a Europe-wide infrastructure of language materials and technology.

»» **Giellatekno**, the centre for Sámi language technology, applies methods of language technology based on an open source code to create digital infrastructure. The centre is creating proofing tools, text processing tools, language learning software, digital dictionaries and text-to-speech applications for Sámi and other northern languages, especially Sámi languages and Kven. However, since Norway is part of the Barents Region, the open documentation and interactive utilisation of application protocols for other languages are also possible.

»» **Helsinki Finnish-State Transducer Technology, HFST**, produces automatic software based on open source code for the morphological recognition of natural languages. Specifying the paradigmatic morphology of a form of language, including headword forms and inflection classes, are integral components in the creation of the software. Essentially, these headword lists, which include inflection information, are the same that are used to create dictionaries. With minor changes, the same automation work can be made use of in proofing applications, text recognition software and morphological parsers.

»» The **META-NORD** project aims to establish an open linguistic infrastructure in the Baltic and Nordic countries.

»» **The Language Bank of Finland** is an aid of linguistic research provided by CSC (IT Center for Science). It consists of an extensive array of text and speech material as well as applications suitable for their study in an efficient computational environment.

»» **The Finno-Ugrian Libraries of Russia** portal is a database that contains links to the webpages of Finno-Ugrian libraries in Russia. The libraries also aim to create an electronic collection that would also contain publications in Finno-Ugrian languages. The project is being carried out in collaboration with the Institute for Russia and Eastern Europe, the Slavonic Library of the University of Helsinki, and the M.A. Castrén Society.

Mission

The objective of the Kone Foundation Language Programme is to advance the documentation and status of small Finno-Ugrian languages, the Finnish language, and minority languages in Finland.

Vision

The Kone Foundation's Language Programme will produce scientifically relevant research and document small Finno-Ugrian languages, the Finnish language, and minority languages in Finland. Both the scientific community and all language users will profit from the results of this research and documentation. Furthermore, research methods based on interaction will be promoted.

Minority Languages of Finland

Swedish*, Finnish and Swedish Sign Language, Northern Sámi, Inari Sámi, Skolt Sámi, Karelian, Romani languages, Tatar and immigrant languages.

* As the second national language, Swedish has a strong status in Finland. Furthermore, Swedish is well researched in Sweden where it is the dominant language. The other minority languages of Finland are therefore given priority in the Language Programme.

Finno-Ugrian Languages

- a) Baltic-Finnic languages: Finnish*, Estonian*, Livonian, Võro, Seto, Votic, Ingerian, Ingerian, Veps, Ludian, Livvi-Karelian, South Karelian, Viena Karelian, Kven, and Meänkieli.
- b) Sámi languages: Southern Sámi, Pite Sámi, Lule Sámi, Ume Sámi, Northern Sámi, Inari Sámi, Skolt Sámi, Kildin Sámi, and Ter Sámi.
- c) Mordvinic languages: Erzya and Moksha.
- d) Mari languages: Hill Mari (Western Mari) and Meadow Mari (Eastern Mari).
- e) Permian languages: Udmurt, Komi-Permyak and Komi-Zyryan.
- f) Ugrian languages: Khanty, Mansi and Hungarian*.
- g) Samoyedic languages: Nganasan, Nenets, Enets, Selkup, Mator** and Kamas**

* In addition to Finnish, Estonian and Hungarian are dominant languages in Estonia and Hungary and their research enjoys a far stronger position than Finno-Ugrian minority languages. Therefore, the Kone Foundation Language Programme prioritizes small Finno-Ugrian languages, in addition to the Finnish language.

** The Mator and the Kamas language can already be considered dead languages. This does not, however, diminish their linguistic significance.

The Mission: Documentation of Small Finno-Ugrian Languages, Finnish, and Minority Languages in Finland

The main objective of the Kone Foundation's language programme is to advance the documentation of small Finno-Ugrian languages, the Finnish language, and minority languages in Finland.

In essence, the documentation of a language includes observing, recording and scientifically analyzing spoken, signed or written phenomena of the language. In the documentation process the recorded material is used and observations regarding it are made. Documentation has also been evolving during the past decades with electronic corpora and language technology applications becoming increasingly popular.

On a practical level, documentation includes both field work and archiving involved in acquiring and maintaining language records as well as research. The research can focus on vocabulary, morphology, syntax and discourse. One of the most thorough forms of documentation is lexicography. In a broader context, documentation can also include the study of a language from a sociolinguistic point of view. In this case, issues such as the position of a language and the kind of linguistic identity sustaining it are under investigation.

The language programme pays particular attention to charting, collecting and maintaining linguistic corpora. Moreover, the material must be refined and its availability and usability must be at the highest level.

Furthermore, it is important to assure that both old and new corpora are made available for the open and interactive use of not only the academic community but of the society as a whole. The Language Programme encourages open interaction between language users and researchers during the various stages of the documentation process, for instance with the methods of citizen science, for example. Similarly, it is fundamental that language users are able to find information on their own language. In other words, the information gained during research must also be returned to the language community. (Please refer to page 17, *Advancing a Culture of Openness and Interaction*)

The Kone Foundation has dedicated itself to advancing interdisciplinarity for many years. Recently, the combination of the perspectives of science

and art has also come to the fore. The projects devised by the Language Programme are also encouraged towards multidisciplinary, for example by making use of different kinds of materials and methods. Interaction between art and scholarly research is also promoted. (Please refer to page 19, *Advancing Multidisciplinarity and Crossing New Boundaries*)

**ENHANCING THE AVAILABILITY, ACCESSIBILITY
AND USABILITY OF CORPORA**

The objective of the Language Programme is to make sure that both old and new corpora are made available for the open and interactive use of both the academic community and the society as a whole. The existing corpora must be charted and old records, tapes, scripts and print products must be rescued. This often involves supporting the microfilming and digitizing of old material. Open source codes and state of the art audio-visual equipment are favored in the recording of new material.

The National Library of Russia in St. Petersburg, for example, contains large amounts of written material in Finno-Ugrian languages. These include textbooks from the beginning of the Soviet Union, educational literature, fiction, magazines, and journals. The Language Programme intends to make this material available for open use. There are also numerous language materials in Finland that still await digitization, such as old recordings of minority languages and recordings that have become obsolete during the development of recording software (the Sign Language interview tapes of The Finnish Association of the Deaf, for example).

There is also an abundance of information collected during the past two hundred years that should be made readily available to a larger audience and returned to the knowledge of the people. In particular, a large amount of material on the people and languages of Finland was collected during the national awakening and enlightenment periods of the 19th and 20th centuries. In addition to their interest in their own language, Finnish scientists, explorers and laymen were also largely involved in studying and collecting information about Finno-Ugrian peoples and languages. It is important to chart and ensure the availability, accessibility and usability of interviews and other language use data obtained through this kind of

fieldwork. This extensive information ought to be returned to the language community so that it can be profited from and freely available to find, obtain and read. This can be achieved for instance with the help of library cooperation as a means to advance not only research but metalinguistic awareness and cultural understanding as well.

In order to make sure that various materials are available to users, search databases require sufficient descriptive metadata files. It would be best to create these kinds of descriptive data files already during the data collection phase.

It is the objective of the Language Programme to not only ensure the availability of language materials, but to make it easy for different users to approach and use them. A prerequisite for the usability of corpora is a digitizing process that uses language recognition, so that they can be used with universally and freely available applications. Understanding the linguistic elements of the material requires the use of electronic dictionaries and machine translation. Language recognition is supported by the availability of lexical and grammatical information as well as complete morphological analysis. The next phase in increasing usability includes both the more extensive recognition of corpora and their contents, and the compilation of broader metadata files. Citizen science, for example, can be used for this purpose to verify the results of automated recognition and description. (Please refer to page 17, *Advancing a Culture of Openness and Interaction*)

It is also important to pay attention to corpus infrastructure. The larger is the portion of material gathered from various sources that end up in the same environment, the larger is the significance of each separate piece of information within the context of the whole. With shared solutions, networking as well as the creation and stabilization of technical platforms suited to different data, it is possible to more efficiently assure that the results of data collection and refinement end up being used by the scientific community and other users. In the future, it is important to build dynamic systems that researchers and other users alike can continuously improve and enrich by further developing open source code systems, for example.

Most material is protected by copyright. In general, copyrights are valid for seventy years after the death of the author. For this reason, ensuring

Taking into Consideration All Kinds of Language Users

The documentation process and the involvement of the language community must also take into consideration the existence of a wide spectrum of language users. In addition to ordinary language users, there are a variety of professionals that work with language, such as teachers, writers, poets, journalists as well as broadcast hosts and commentators. All language users have a preconception of the language they use and of their own roles as users of that language. Language professionals often have a standardizing role. In the case of languages that do not have a coherent standard form, the role of a standardizer may even be significant. Problems arise if the standardizers have different notions of the standard form. An incoherent standard form gives opportunity for variation, which can be both a good and a bad thing. The positive side is that it allows language users more freedom in their expression; the negative aspects include conceptual ambiguity and a lack of conciseness.

A coherent language is the result of collaboration between several groups of language users. Eventually, once the language is seen as stemming from a healthy foundation and is free of pressure, it is respected as the norm. In order to achieve such circumstances, it is necessary to commit to long-term cooperation during which the contributing parties treat each other as equals and with due respect. Expert help must also be valued. At its best, the different changes that occur in a language happen gradually over time and without any large upheavals. As a result, the changes are more easily acquired and language users are not estranged from their own language.

the open use of many materials can be extremely complicated or even impossible. In collecting new material, it is important to pay particular attention to contracts and licenses related to copyrights so that the ideals of open and free use can be carried out to the highest possible degree. The possibility of attaining material for noncommercial use can be determined with the help of Creative Commons licenses, for example. The holder of a Creative Commons license can choose to waive some of the rights that are the default settings of copyright.

ADVANCING A CULTURE OF OPENNESS AND INTERACTION

Another key objective of the Language Programme is to foster a culture of openness and interaction in linguistic research. This entails, for example,

the unlimited availability, accessibility and usability of source material and research results via a virtual library, for instance, as well as the participation of the language community in various stages of documentation and application of research results.

In order for the documentation and research results to better serve the public, the collected information must be reworked into a more intelligible and interesting form. The Language Programme supports the participation of the language community in creating new material and enhancing the usability of old material by applying the methods of citizen science, for example. Community participation and interaction are also supported by projects that combine art and research.

Citizen science is a form of interactive research that includes the participation of researchers, students and any interested citizens. Citizen science is based on the work of trustworthy volunteers, who help in observation, measuring and calculation work. Citizen science is a way of obtaining new material and carrying out large-scale proofing. An early example is the lexicography of the Dictionary of Finnish Dialects, which handsomely illustrates the help offered by responsible volunteers during the compilation of the work.

Interactive research can also benefit from crowdsourcing i.e. collaborating with an indeterminate group to carry out development and proofing work, for example. A excellent good example of crowdsourcing is an initiative by the National Library of Finland, Digitalkoot (Digital Volunteers), which uses two online games to help digitize archival material: the answers of the players are used to evaluate and correct the quality of automated character recognition.

The difference between citizen science and crowdsourcing is not necessarily very pronounced, but the participants of crowdsourcing tend to be unknown, whereas “citizen scientists” refer to specific people collaborating with researchers and applying research methods.

An essential aspect of citizen science and crowdsourcing is that the efforts of volunteers are also made available for anyone to use and to develop. In fact, many small Finno-Ugrian languages do already have their own Wikipedia and user interfaces for Skype, Vkontakte and OpenOffice, which are all the result of crowdsourcing. However, there are fewer vol-

unteers developing software for small languages than there are for larger languages and so added support is most likely necessary.

Using the help of native speakers for linguistic research is also an excellent approach. As interviewers and information gatherers, native speakers work best within their own extended family or village. They are both insiders and members of the scientific community and as such they are able to determine the validity of acquired materials themselves and also inspire confidence in their interviewees. Furthermore, as members of their communities, their presence can cause an interviewee to speak less artificially or unnaturally. In addition to being useful in linguistic research, the efforts of native speakers can also strengthen the status of the language and the linguistic identity as well.

Civil organizations and their networks can be used for the purpose of recruiting and mobilizing volunteers for citizen science and crowdsourcing. It may be especially fruitful to acquaint young volunteers with citizen science and crowdsourcing as they may already possess the skills needed to use computer software.

ADVANCING MULTIDISCIPLINARITY AND CROSSING NEW BOUNDARIES

Both multidisciplinary and interdisciplinarity can bring new perspectives, opportunities and solutions to research. Linguistic material can also be used in other fields and corpora can even serve as a valuable nexus of multidisciplinary. Linguistic corpora are useful not only for linguistics but for historical and sociological research as well. Similarly, musicological or medical corpora can also offer valuable material for linguistic research, if the multidisciplinary potential of the corpora is put to use.

Shared corpora may be a fruitful starting point for cooperation between different fields. However, linguistics, social sciences and historical research can benefit from collaboration in other ways as well. Moreover, language technology applications have been developed for different disciplines: criminology and medicine both use language technology for voice recognition, for example.

In addition to encouraging multidisciplinary, the Language Programme supports projects that combine the perspectives of art and sci-

ence. For some time now, interdisciplinarity in both science and art has given rise to new questions regarding their respective subject matter they are involved with and art, and has enriched the way they are studied. The Language Programme demonstrates the desire to encourage the use of perspectives that combine both science and art, and the breaking of boundaries between different disciplines.

HOW ARE THE OBJECTIVES ACHIEVED?

Funding and Initiating Projects

The Language Programme is taken into account in all the activities of the Kone Foundation. In order to achieve its objectives, the Foundation will arrange the allocation of specially allocated grants and be involved, to a varying degree, in the realization of new projects. In addition, the Language Programme will be incorporated in the existing grant application rounds as the selection process will focus specifically on projects that fulfill the objectives of the Programme. Relevant parties may also be prompted to apply for grants during the Foundation's grant application rounds.

So far, the Kone Foundation has not been an active party in the projects it funds. The Foundation has, however, arranged seminars together with the funded parties. Nevertheless, the Foundation is not involved with the contents of the research. In addition to affording traditional grants, the Foundation is an active participant in certain aspects of the Language Programme, and is also involved in planning projects and in defining their content.

Advancing Cooperation and Good Practice

Furthermore, in order to achieve its goals, the Language Programme will also make use of existing networks, advance networking in various ways and encourage cooperation between people working within the same field. (Please refer to page 17, *Taking into Consideration All Kinds of Language Users*)

The Language Programme contains various aspects that will be advanced in cooperation with organizations and projects that already exist in

the field. This is particularly relevant as regards copyright and infrastructure as several different digitizing and language technology development projects have already taken measures for this same purpose. Tools that will facilitate communication and cooperation between different projects are also being developed within the framework of the Language Programme.

Various projects concerned with digitizing Finnish language materials or using language technology in some other way have been and are presently being carried out in Finland. From the point of view of the Language Programme, many such initiatives can be considered pilot projects that can possibly later be applied to the study and documentation of other languages. Therefore, the Language Programme intends to support these types of projects and applications. Good practices displaying openness and interaction are also particularly worthy of support.

The concrete measures i.e. specific projects and specially allocated grants based on the objectives of the Language Programme are outlined in the following section *Measures 2012–2016*.

Concrete Measures 2012–2016

This section on the content of the programme will be updated and elaborated each year and will also take into consideration any later requirements the projects may have. More information on the Language Programme and the funded projects is available on the Foundation website.

Approximately six million Euros have been reserved as an overall funding during 2012–2016.

COMMUNICATION AND COOPERATION BETWEEN PROJECTS

The Foundation will arrange seminars and workshops on a yearly basis to support the communication and cooperation between projects. Various web-based solutions designed to facilitate communication and cooperation are also being evaluated. Furthermore, various ways of utilizing and profiting from citizen science and crowdsourcing are also being considered.

In planning the concrete measures to be taken in the Language Programme, it is important to consider the timing of technical solutions in relation to work done on the content of funded projects. The implementation of corpora, for example, can easily form bottlenecks. These types of issues will also be considered at the workshops.

PROJECTS FUNDED THROUGH ANNUAL GRANT APPLICATION ROUNDS, SPECIALLY ALLOCATED GRANT CALLS, AND PROJECTS INITIATED BY THE FOUNDATION

Projects Funded Through the Annual Grant Application Rounds of the Foundation (2012–2016)

The Foundation's annual grant application rounds will take the Language Programme and its objectives into consideration. The Foundation may also prompt relevant parties to apply for grants to support their projects. The focal points of the Language Programme were already taken into consideration in the grant allocation decisions of 2011.

»» The compilation and translation of dictionaries, textbooks, other study material and other books compatible with the Language Pro-

gramme are supported by the *non-fiction grants* of the Foundation. In 2011, funding was granted to the *Dictionary of Old Literary Finnish*, which had been temporarily discontinued due to cuts caused by the statal productivity programme. The work continues at Kotus with the support of the Kone Foundation.

»» *Saari Manor residencies and arts and culture grants* support artists from minorities. The Foundation wishes to support the writers and translators of kindred peoples in particular, and to welcome them to the artist residence at Saari Manor. Other arts and culture projects relevant to the objectives of the Language Programme are also supported. For example, Professor Jyrki Nummi's project, *the Finnish Literature Bank*, was granted funding by the Kone Foundation in 2011. The plan is to design and test a digital collection of classical Finnish literature as well as an operational model for a user-driven online service for the collection. »» The focal points of the Language Programme are taken into account in the allocation of annual *researcher grants and project grants*. In 2011, funding was granted, for example, to a project directed by Tiina Onikki-Rantajääskö entitled *the Cognitive Grammar of Finnish in the Framework of Finno-Ugrian Languages*.

*A Project Started by the Foundation: The Digitization of the
Finno-Ugrian materials at St. Petersburg, Pilot Project (2012–2013)
and a Possible Follow-up Project (2013–2016)*

Please refer to page 15, *Enhancing the Availability, Accessibility and Usability of Corpora*

The Kone Foundation has awarded a grant to the National Library of Finland for the digitization of the Finno-Ugrian materials in the National Library of Russia in St. Petersburg. As an initial pilot project several Mari and Mordvinic newspapers from the 1920s and 1930s as well as a selection of Veps and Ingrian books were chosen to be digitized. It is important to take into consideration the needs of the researchers and to have them participate in the planning and execution of the project from the very beginning.

The follow-up project may include collaboration with the national libraries of other constituent entities of Russia. In the summer of 2012, the

annual conference of IFLA (The International Federation of Library Associations and Institutions) is being held in Helsinki, which provides an apt opportunity for a project planning meeting.

*A Project Started by the Foundation: Morphological parsers
for Finno-Ugrian minority languages (2012–2016)*

The Foundation is funding a project that, as a result of cooperation between language technology and linguistics, will produce text parsers (transducers) based on an open source code. These parsers can be used for proofing programs, the analysis of text corpora and the recognition of scanned texts. The preliminary plan is to have one language technologist working on the project, who will coordinate with four other linguists focusing on different languages. The project could include languages such as Mari, Moksha, Nenets and Veps. The project will be carried out at the Department of Modern Languages at the University of Helsinki. The stages of the project are:

- a) A basic description of morphophonological and regular inflections for the parser
- b) A description of morphophonological and regular inflections with a list of headwords including inflection information (approximately 20,000 words).
- c) A description of morphophonological and regular inflections with a list of headwords including inflection information as well as preliminary glossing and emphasis based on frequency.
- d) The development of applications that advance lexicography and text recognition.

For some languages there already exist at least preliminary inflectional and morphophonological descriptions or glossaries, which will help in the beginning. Descriptions made of one language can be used as models to describe a kindred language. All Finno-Ugrian language groups, apart from Samoyed languages, already possess at least one preliminary description of their morphophonology and inflection.

Close cooperation with the CLARIN project is crucial in order to avoid overlapping and diverging solutions.

A Project Started by the Foundation: Collaboration between Finno-Ugrian theatre practitioners and linguists (planning 2012, execution 2013-2014)

Preparations for the project will begin in 2012 but the actual funded project will be carried out in 2013-2014.

Specially allocated grant call: Corpora (2013, beginning of the year)

The focus of the corpus project already started by the Foundation lies particularly on the Finno-Ugrian material at St. Petersburg. The corpora play a significant role in the Language Programme, and other projects related to the subject may also be funded. The special grants can be directed towards projects involving the rescue of language material, their beneficial utilization, and citizen science. Particular attention will be paid to the material collected by Finnish researchers during the late 19th and early 20th centuries. This corpora-related specially allocated grant may also be used for a part of the St. Petersburg digitizing follow-up project.

Specially allocated grant call: Lexicological and lexicographic research and vocabulary work

Lexicographic and lexicological research ought to work more closely together than they presently do. Their collaboration is essential in the recording and analysis of vocabularies and in communicating the content of the vocabularies to common knowledge.

The specially allocated grant dedicated to lexicography and lexicology ought to be used at the beginning of the Programme so that vocabulary sources can be added to headword lists. Subject matters can include the formation of terminology in Finno-Ugrian languages, and the 1920s and 1930s lexicon of Finno-Ugrian languages, for example. Not only is it important to develop terminology in minority languages, but the special terminology of dominant languages must also be maintained and updated.

The results will benefit the development of both *The Bank of Finnish Terminology in Arts and Sciences* and the lexicon of small Finno-Ugrian languages.

*Specially allocated grant call: The compilation of grammar books
and research on the central aspects of grammar*

There is a lot of work to be done regarding the grammar of Finno-Ugrian minority languages: even the languages for which grammar books exist require clarification in many areas. Most of the existing grammars are extremely narrow generalizations, often based upon small and one-sided material.

Minority language grammars could be compiled as doctoral theses. Such a thesis could, for example, analyze and document the morphophony, regular inflection and derivation necessary for a morphological parser of a given language as well as develop word form recognition with the help of syntax information.

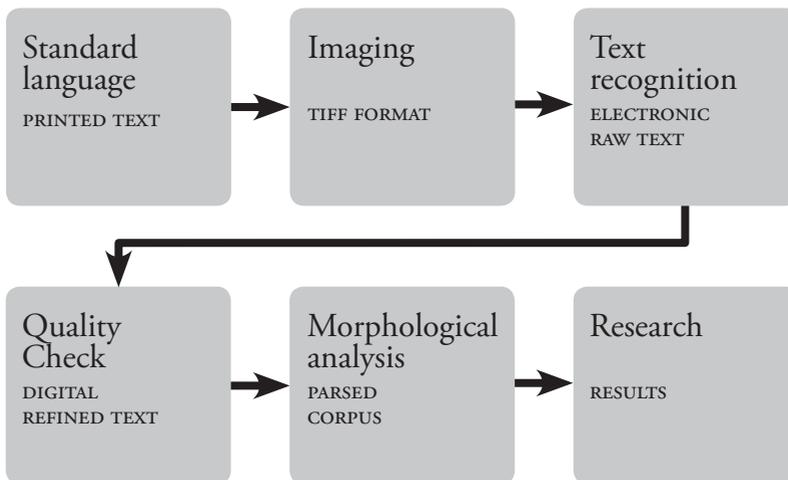
*Specially allocated grant call: Language revitalization
and development research (2015)*

Projects related to language revitalization and development research can be related, for instance, to problems associated with urbanization.

An Example of the Use and Development of Language Technology Applications

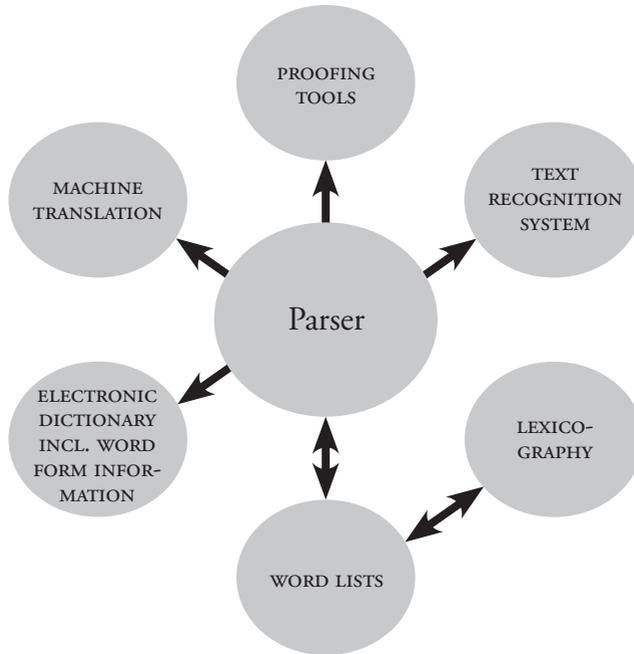
The possibility of cooperation between the different projects will be taken into consideration during the brainstorming and funding phases. The use and development of various language technology applications during different projects or during the different stages of the same project serve as an example of the possibilities created by the cooperation and synergy between projects (see the Figures).

From Printed Text to the Utilization of Parsed Text



The documentation of written text begins with a text in printed or digital form. A printed text must first be scanned and recognized as a text. The quality of the recognized text depends on the structure of the recognition software. A simple application will recognize letters and punctuation marks. However, the quality of printed text is not always optimal as letters may be faded, for instance. In such a case, the quality of character recognition can be enhanced with word form recognition, which is possible with morphological parsers.

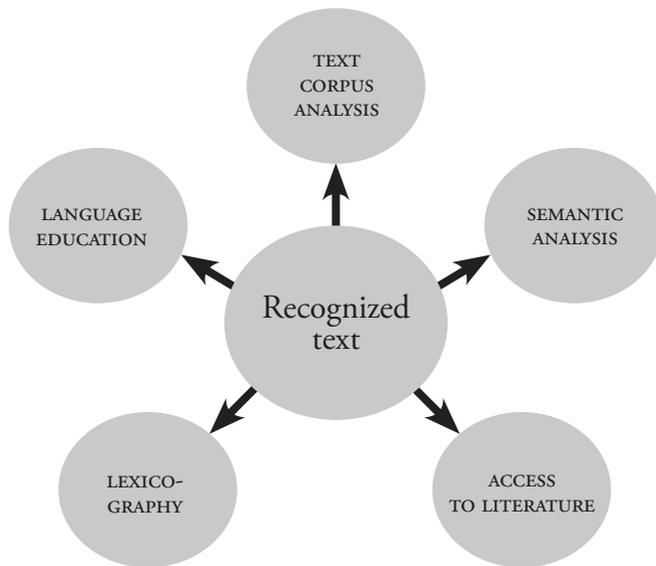
How are Parsers Applied?



A parser is an automatic application that combines headword lists and inflection information. It can be used for both character recognition and spell check programs, dictionaries and machine translation.

The headword lists necessary in parsers can also be used for lexicography. By using a shared database, semantic information can be added to parsers, which, in turn, will help to make the written texts of other languages easier to understand. Headword lists are expanded with unrecognized words. Existing digital texts and revised corpora serve as experimental material that can be used in the development of parsers. As recognition ability increases, the accuracy of applications using parsers also grows.

Applications of Recognized Texts



As the amount of recognized text increases, access to literature is also improved. The texts can be used as corpora or in language and culture education, and they can also be used for semantic research, for instance lexicography.

Monitoring the Programme

Even though the Language Programme will not have a steering group, both Finnish and foreign participants will be informed of the progress and the Kone Foundation will be kept up to date on other related activities. This will include participation in the Finnish Culture Foundation's kindred peoples' workgroup meetings and active reporting on projects.

Methods for monitoring the Programme and its efficiency will be considered during the first year of its operation.





The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that every sale, purchase, and payment must be properly documented to ensure the integrity of the financial statements. This includes recording the date, amount, and purpose of each transaction.

Secondly, the document highlights the need for regular reconciliation of bank accounts and credit cards. By comparing the company's records with the bank statements, any discrepancies can be identified and corrected promptly. This helps to prevent errors and ensures that the financial data is up-to-date and accurate.

Another key aspect is the proper classification of expenses. It is crucial to categorize each expense correctly according to the company's chart of accounts. This allows for a clear understanding of where the money is being spent and facilitates the preparation of financial reports.

Finally, the document stresses the importance of timely reporting. Financial statements should be prepared and reviewed on a regular basis, typically at the end of each month. This provides management with the information they need to make informed decisions about the company's operations and financial health.

In conclusion, maintaining accurate and up-to-date financial records is essential for the success of any business. By following the principles outlined in this document, companies can ensure that their financial data is reliable and that they are in a position to make sound financial decisions.

KONE FOUNDATION
Tehtaankatu 21 B 49
00150 Helsinki
Finland
koneensaatio@koneensaatio.fi
Tel. (09) 260 0617
Fax (09) 260 0618
www.koneensaatio.fi