New technologies are bringing about wide-scale changes. They have enabled the creation of a global information space, global communications and global commerce. These are trends which are of serious concern to smaller nations. Will their languages be heard in the global village? Could their cultural heritages become nothing more than an archaic legacy? This is an issue which is being discussed by specialists at the Baltic and European level alike.

Latvia's national informatics development program has this to say on the matter: "Serious, even overacted attention must be devoted to supporting the distinct cultural environment of Latvia, especially the Latvian language, in the process of creating the Information Society. The very existence and future development of the Latvian nation, its culture and its language depends on this issue being addressed properly."

The Council of Europe, the European Union and UNESCO have declared 2001 to be the year of European languages. Among other language protection and development tasks, the issue of using languages in cyberspace and in new technologies has arisen. UNESCO is preparing a recommendations concerning the promotion and use of multilingualism and universal access to cyberspace. The draft document includes several important aspects of maintaining and developing small languages and cultures within the Information Society. Here are a few of the requirements which UNESCO is defining:

- The promotion, respect and use of all languages in cyberspace must be reaffirmed;
- Language diversity in cyberspace must be broadened by creating content in all languages that are widely used internationally, as well as in other languages at the regional, national and local levels;
- All forms of linguistic segregation in terms of access to cultural and scientific information and knowledge must be prevented;
- Global academic efforts to develop automated translation services

must be supported;
- Strong national policies on the crucial issue of language survival must be formulated.

The 5th Research, Technology Development and Demonstration Framework Program (FP5) of the European Union includes the key action “Multimedia content and tools”, which aims “to enable linguistic and cultural diversity and contribute to the valorization and exploitation of Europe’s cultural patrimony, to stimulate creativity, and to enhance education and training systems for lifelong learning.”

THE BALTIC STATES AND LANGUAGE TECHNOLOGIES

This work is particularly complicated for the Baltic languages. The bulk of technological support for most European languages has already been developed. FP5 projects mostly focus on improving these technologies and on various practical ways in applying them to existing products. These tools are beginning to play their part in achieving the European Union’s objective of setting up a user-friendly Information Society. Basic technologies such text analysis tools, machine translation tools, speech technologies and so-

Tilde’s electronic dictionary provides translation from and into Baltic languages.
Tilde provides Internet TV services in partnership with leading Baltic content providers.

The fact that makes this particularly difficult is our inability to participate in FP5 projects on equal terms, because the technologies that we are developing are of no interest to potential European partners. The comparatively small Baltic market means that the international corporate sector has only limited interest in technological solutions for our languages.

This means that we have to do the work on our own, taking full advantage of cooperation that is available at the Baltic scale. One of the most recent examples of such cooperation was a seminar which Tartu University staged in December 2000 – “Estonian-Latvian Language Technology at the Turn of the Century”.

There are wider possibilities in the cultural sphere. We can take advantage of the experience of developed countries, and we can participate in FP5 and UNESCO projects. Here results will largely depend on the interest, initiative and energy of the people who are involved.

**TILDE – SOLUTIONS FOR EBALTICS**

The company mission for Tilde is to develop solutions for Baltic users in areas that are based on our languages, culture, business environment and local information and entertainment needs. Tilde is a group of companies which cover the three Baltic States, which means that it can overcome the restrictions that are imposed by small local markets. The fact that the Latvian and Lithuanian languages are related makes easier the development of language tools for those languages.

Tilde has formulated its activities as “Solutions for e-Baltics”, and there are four major areas of activity:

- **eLanguage**: Tilde is developing software solutions for the Baltic languages which offer the same possibilities that do similar tools for major languages. Software providing full solutions for Latvian and Lithuanian language needs – Tildes Birojs and Tildes Biruas – is being used by tens of thousands of people. The Tilde Localization Department localizes and adjusts IT products which are developed by international software companies for Baltic users;

- **eCulture**: Tilde is working on the creation of new possibilities for the cultural riches that exist in the digital world;

- **eManagement**: Tilde builds business solutions that are adapted to local legislation and needs. More than 1,300 users are working in the Tilde Jumis software environment. The company also creates information management solutions for enterprises

on the basis of portal technologies;

- **eMedia**: Tilde has created Internet Television which allows users in all three Baltic countries to watch local TV broadcasts and tailor-made videos on the Internet (TV.LT, TV.LV and TV.EE).

**MT – A REMARKABLE STEP ON THE GLOBALIZATION PATH OF THE BALTIC LANGUAGES**

The idea of providing machine translation possibilities between various languages is extremely significant. If implemented, it could provide for the equality of all languages, enabling people to obtain the necessary information in their native language. Tilde is encountering the desire of people to use machine translation tools for the Baltic languages more and more often. An up-to-date user has discovered this functionality for major languages such as English, Spanish, German and French, and wants it for Latvian, Lithuanian and Estonian, as well.

It must be understood, however, that MT systems currently have certain limitations. They can be used to understand the idea of a text, but they cannot be considered a tool for obtaining literary translations. That is why MT is mostly used for translating Internet resources, E-mail messages and various draft documents for the European Union.

MT tools will remain imperfect for many years to come, but they’re already very useful within certain limited spheres – retrieval of practical information from several European languages at this time. When one is planning a journey, for example, one can obtain transport timetables from Portuguese Web pages that are automatically translated into English.

Tilde considers the development of MT tools for the Baltic languages to be one of its core objectives. The first stage of the project is to develop an MT system for English-Latvian translation. Considering the relative closeness of the Latvian and Lithuanian languages, these technologies can be enhanced to include Lithuanian, as well.

Tilde plans to develop the MT system gradually. There are several reasons for this. First, the system will be developed from several mutually compatible modules, each of which per-
forms a certain language processing task and can be used independently of the MT system. The word-form generation/recognition module is used within spell-checking tools, and it helps to determine the agreement of various words in a sentence.

Second, the development of this kind of system is a complicated and labor-intensive process. The plan is to build up the system from the following components:

- The morphology component, which ensures the recognition/synthesis of word forms;
- The syntax component, which performs syntactic analysis and synthesis;
- The semantic component, which provides a description of semantic meaning and makes it possible to generate a language-independent representation;
- The lexicon component, which will include morphological, syntactic and semantic information about each word;
- The translation memory, which will store text fragments that have already been translated, thus avoiding the need for repeated translations and securing translations of a higher quality;
- The knowledge base, which will contain facts about the world and information about each word’s meaning.

### SPEECH TECHNOLOGIES FOR THE BALTIC STATES

Speech technologies for speech synthesis and recognition are rapidly being developed for the world’s major languages. These technologies can recognize human speech and generate speech from text. Speech technologies have received particular attention during the last several years, along with the rapid development of mobile phones and small computing devices. Speech technologies for the English language are being included in widely used software products such as Microsoft Office XP.

Tilde plans to develop text-to-speech technologies for text that has been prepared electronically. These can be particularly useful when working with mobile phones, small mobile devices and car-integrated computer systems.

Text-to-speech is not that important for desktop users, but it can be very useful in specific situations such as comparing entered information with printed text. Speech systems are also irreplaceable for people with sight disabilities.

The wider use of speech synthesis requires software which can pronounce text with human-like expression. The development of a text-to-speech system includes the following requirements:

- Development of a language lexeme database;
- Analysis of word forms with morphology tools;
- Development of rule-based algorithms to transform text into phonetic transcription;
- Preparation of speech synthesis rules;
- Development of rule-based algorithms which describe the intonation within a sentence;
- Development of software for text analysis and sound synthesis on the basis of algorithms.

Speech recognition is more complicated than speech synthesis. It is widely used in dictation systems which allow for the automatic transformation of spoken text into written form in various information search tools and even automatic booking systems.

When language systems are developed, it is of key importance to ensure cooperation between academic and industrial institutions. Significant research in the field of language technologies, including speech technologies, is taking place at the Artificial Intelligence Laboratory of the University of Latvia’s Institute of Mathematics and Computer Science under the supervision of Dr. Andrejs Spektors. The programs of the Latvian Science Council are also opening up possibilities for academic-industry partnerships.

Speech synthesis and speech recognition are also being researched in the other Baltic countries. Research activities in Lithuania have mostly been centered around speech synthesis at the Department of Informatics of Vilnius University. In Estonia, the Institute of the Estonian Language in Tallinn is working on speech synthesis, while the Laboratory of Phonetics and Speech Technology at the Institute of Cybernetics of the Tallinn Technical University is working on speech recognition.

Optical character recognition is another language technology-related sphere. The international software developers which are working in this field – Abby, for example – have successfully created applications which allow hard copy text to be turned into electronic format. This is software which also works for the Baltic languages.

### LOCALIZATION FOR THE BALTICS

The “native” language for most IT systems is English. If users are profes-

![Thesaurus: Latvian](image)

Language tools provided by Tilde include proofing tools, a hyphenator and a thesaurus.
professionals, they understand English interfaces and manuals well enough. Present-day technologies, however, have turned something that used to be the privilege of the few into a daily routine for the many. For numerous users, the language gap severely complicates technology use. Being aware of this fact, many international corporations are adjusting their products for Baltic users.

Tilde is the only company in the Baltics which offers professional IT-related localization services for all three Baltic languages. At one time this caused a great deal of incredulity among professionals, who claimed that the localization of widely used software is too expensive and that our market is too small for manufacturers to invest in this process. Professional attitudes in terms of localization and the use of state-of-the-art translating tools, however, have enabled Tilde to provide quality services at a price that is acceptable for international corporations.

The localization process involves several stages of work:

• Preparation of a product-specific glossary;
• Translation of user interfaces;
• Translation of help functions and other user assistance information;
• Adaptation of products to local cultural and legislative requirements;
• Compilation of the product;
• Testing and fine-tuning.

Advanced translating software dramatically reduces the workload of translating, improving quality and consistency. The software stores the whole of a translated material in its memory, proposing phrases and sentences already translated on the basis of text analysis. As the vocabulary used in software products is relatively limited, and a considerable part of the text material is retained in various versions, there are significant economies of time and labor. The specialized and universal dictionaries which Tilde has developed provide consistent usage of the latest terminology during the localization process.

The ongoing localization projects at Tilde include the Client Access part of the IBM AS/400, Navision Financial Software, interfaces for Ericsson and Nokia mobile phones, and various Microsoft materials, to name but a few.

**ECULTURE – FROM CULTURAL HERITAGE TO NEW DIGITAL CONTENT**

The cultural heritage of the Baltic countries should be transferred to new Internet technologies, and new cultural values should be created. Tilde released Latvia’s first multimedia product for popular usage – the CD-ROM encyclopedia “The History of Latvia” – in 1998. A new version of the encyclopedia is being produced in English, with more than 700 entries, more than 1,000 pictures, 70 video and audio clips, and diverse multimedia galleries which provide a wide and multiform insight into the complex history of our country.

The publishing house “Nacionālais Apgāds” is partnering with Tilde for a new and larger project – the Latvian National Encyclopedia. It will be available in printed form, on CD-ROM and on the Internet. The first product, an encyclopedic dictionary with 70,000 entries, should be ready at the end of 2001. Subsequently we will compile 130,000 advanced entries, which will take up 20 volumes in printed form. Project manager Pēteris Apinis is convinced that a very comprehensive information product can be created in Latvian and that it will be used for decades. Tilde is also participating in a project called “Dainu skapis” (Cabinet of Folk Songs) which is being run by UNESCO’s Latvian Committee under the leadership of the director of the Latvian National Library, Andris Vīlks. The objective of this project is to participate in UNESCO’s “Memory of the World” program and to make Latvia’s cultural heritage available on the Internet.

**A CD-ROM TO PROMOTE CULTURAL VALUES**

An example of how cultural values can be rendered in digital form is a unique CD-ROM called “Juris Podnieks: The 20th Century as Seen by the Latvian Filmmaker”. It was released by the International Center of Cinema in cooperation with Tilde to honor the 50th anniversary of the birth of the late Juris Podnieks, who was an outstanding Latvian documentary filmmaker. The CD-ROM contains Podnieks’ biography and filmography, as well as fragments of and annotations to his various films. There are also reviews by film critics and memories of the filmmaker from his contemporaries. The CD-ROM provides a deeper insight into the history, environment and culture of Latvia.

The 100th anniversary of another outstanding figure in Latvian culture, the poet Aleksandra Čaks, will be observed this autumn. A tribute CD-ROM is planned, and project organizers Ilze Konste, Iveta Ruskule and Jānis Mitrevics are currently looking for sponsors.

The development of advanced language technologies, the digitalization of cultural heritage and the creation of new information resources – all of this will mark a significant contribution from the Baltic countries in the global Information Society.

**REFERENCES**


Vasiljevs, A. “SIA Tilde programmu nodrošinājums nacionālās valodas atbalstam” (SIA Tilde programs which support the national language”, materials from the seminar “Language and Technology in Europe 2000: The Baltic Perspective”, Riga, 10-11 November, 1994, p. 32.