

DICTIONARIES FOR EVERYBODY

Unifob AKSIS makes web-based dictionaries for immigrants. The LEXIN project secured Rosing's Language Prize.

The interactive image on the computer monitor shows a supermarket surrounded by various edible products. A click on the oblong food item with a crust tells you that what you call *rooti* in your Somali mother tongue is called *brød* in Norwegian.

«Many dictionaries are made available online without making full use of the opportunities offered by digital distribution,» says Sindre Sørensen, a developer with AKSIS. «The available space is limited in printed books, but pixels are free. Pictures and illustrations can therefore be used in a totally different way on the internet. And, not least, hyperlinks can be used to link the images to dictionary entries.»

Unifob's Centre for Culture, Language and Information Technology (AKSIS) has been making low-threshold dictionaries for immigrants on behalf of the Directorate for Education and Training for years. So far, web-based LEXIN dictionaries are available in Arabic, English, Kurmanji, Sorani, Persian, Somali, Tamil, Tigrinya and Turkish, using both the Bokmål and Nynorsk varieties of Norwegian as source languages. Four of these dictionaries are also available in printed versions.

«There are probably about 150 languages spoken in Norway, but many of them are used by so few people that it is impossible for commercial players to invest in them. That means that the Norwegian state must take responsibility,» says Tove Bjørneset, who is project manager for the LEXIN project. «Since we use a static source language, it is also relatively quick and cheap to make new dictionaries for other languages.»

At the moment, dictionaries for Thai and Urdu are being completed, while Russian and Polish are in the start-up stage. In addition, work is also underway on developing a search interface for mobile phones, among other things. Simplicity and user-friendliness are key concepts throughout the process, and in 2008 LEXIN was awarded the Rosing's Language Prize for these very characteristics. The jury for the prize, which is awarded by the Norwegian Computer Society and the Norwegian Language Council, praised LEXIN for its plain, fast and practical service of great value to its users.



Project Manager Tove Bjørneset and Software Developer Sindre Sørensen

«We maintain a continuous dialogue with our users and listen to their feedback,» says Ms Bjørneset. «Our launching of *Nynorsk* as a source language for all our dictionaries last year was not a result of guidelines from the authorities, but of the demand from immigrants living in areas where *Nynorsk* is used. This shows the value of the language, both in terms of our identity and as a common cultural denominator.»

Own tools

The idea of a state dictionary project for immigrants was originally Swedish, and is more than 30 years old. The Swedish solution became so popular that both Norway, Denmark and Iceland accepted the offer of borrowing the material to develop corresponding language resources. After a successful pilot project lasting from 1996 to 1999 in which Swedish source language materials and Swedish translations were imported and reused, the first online dictionaries were launched in 2002. The Norwegian-English and the two monolingual Norwegian online dictionaries consist of 36,000 entries, compounds and derivations, while the other dictionaries contain between 7,600 and 11,000.

«LEXIN is a resource that helps its users to achieve a level of language proficiency that enables them to understand and be understood better. In the long run, they will be able to use other, larger dictionaries, which are essential in developing a good mastery of the Norwegian language,» says Ms Bjørneset.

AKSIS has developed the editing tool themselves, and it allows the translators to edit the dictionary material via a special website. This means that the translators can work on LEXIN from anywhere in the world. The Ministry of Foreign Affairs also uses tools from AKSIS, for example in connection with their terminology database for EEA and EU terms, *EØS-EU-basen*.

«Multidisciplinary project groups give us philologists a good opportunity to provide input and communicate our wishes to the developers and programmers, and vice versa,» says Ms Bjørneset. «This continuous collaboration on developing and testing tailor-made tools and resources often results in ground-breaking solutions. It is also an important element in creating job satisfaction in our hectic day-to-day work. Our experience shows that there are many outside of our academic environment who have a great need for our multi-disciplinary expertise without realising it.»



FACTS

Unifob AKSIS does research and development in the following areas:

Language technology. Development of online dictionaries and solutions for language processing. We build language resources and infrastructure for research and development and we offer tools for term extraction, grammar development, annotation and corpus work.

Language testing. We develop and carry out language tests in foreign languages. We conduct research on assessment methods (assessment of, assessment as and assessment for learning), language testing and standard setting.

ePublishing. Research in semantic technology in order to facilitate efficient search in text collections. We develop infrastructure for the publishing of text collections and host digital versions of cultural resources including the works of Ludwig Wittgenstein.

Digital media and eLearning. Research on the pedagogical potential and limitations of digital media and eLearning systems. We analyze social and cultural aspects of new media and offer technology for development and assessment of eLearning systems.

Highlights

- Unifob AKSIS contributed to the publication of famous persons' writings: Henrik Ibsen's poems are part of the project Henrik Ibsen's writings (publishing 2009). On eMunch.no, the publishing of Edward Munch's writings began. 5000 pages of Ludwig Wittgenstein's original manuscripts are being published on wittgensteinsource.org.
- LFG-Parserbanker – tool creating databases of grammatically analyzed sentences was completed and licensed to several companies and universities.
- The LEXIN dictionaries for minority languages won Rosing's language award. These dictionaries have been developed by Unifob AKSIS since 1996.
- The government proposition about Norwegian language (Mål og mening) was received very positively by the research communities. Its focus on terminology and the creation of a Norwegian language bank is especially interesting for us.
- Eli Hagen took over as research director 1 October. She came to Unifob AKSIS from BMW's research and development centre in Munich where she developed speech applications and was program manager R&D in the interior division.

Important projects

The goal of the *Semi-SemAntic Models for cross-sector portals* (SeSAM4-project) is to simplify the usage of semantic web technology and make it more accessible for small and medium enterprises.

The EU-project *Common LAnguage Resources and technology Infrastructure* (CLARIN) shall develop a new electronic research infrastructure for language resources and make these accessible for the human and social sciences.

In the EU-project *Science Created by You* (SCY), we conduct research on how digital teaching aids can make science more interesting

In the project *Termportalen*, we establish a national infrastructure for Norwegian terminology.

Collaborators

Unifob AKSIS cooperates with several universities – especially the human and social science departments at the University of Bergen. Other partners and customers include the national research council, EU, The Norwegian Language Council, businesses and government organizations.

