A boost for the digital infrastructure for linguistics: the Taalportaal

Ton van der Wouden
Meertens Institute
Ton.van.der.wouden@meertens.knaw.nl

As of January 1st, 2016, the first version of the Taalportaal (Language Portal) is on-line (www.taalportaal.org).

The Taalportaal can be described as a new type of comprehensive scientific grammar for Dutch and Frisian, the two official languages of the Netherlands. Written in English, the "lingua franca" of contemporary science, it covers the syntax, the morphology and the phonology of the two languages in a parallel way. Open to everyone with internet access, it is a new type of digital information source for linguists as well as for all other researchers interested in language.

Next to that, the Taalportaal also functions as a hub in the scientific digital infrastructure in at least two ways. Firstly, the grammatical information in the portal is enriched with links to existing other types of digital language resources via the CLARIN infrastructure. For instance, phonological descriptions are illustrated with sound samples from spoken language corpora, morphological descriptions are connected to lexical databases, etymological databases, on-line dictionaries, etc., and syntactic descriptions and examples are enriched with queries into text corpora and treebanks, opening new possibilities for efficient language-related research in the Benelux and all over the world.

Secondly, the Taalportaal's open architecture makes it relatively easy to further expand the content and usability of the Taalportaal. At the moment, a third section of the grammatical database, on the Afrikaans language, is well under way. A following extension will deal with Dutch again, but this part of the Taalportaal will be aimed at another audience, to wit, language professionals such as teachers and writers. This part will be written in Dutch, but as the same architecture will be used, many of the original examples and links can be re-used. Funding is currently being sought for other types of extensions: one may think of new languages being added or connected, but also of adding new types of information or links for the languages already covered: semantic information, historical corpora, and databases on dialectology are only three of the infinitely many possibilities.

In our presentation, we will discuss the underlying (XML-DITA) architecture of the Taalportaal, show some of its usage potential for various types of users, and reflect on possible extensions.