NTNU
Innovation and Creativity

TypeCraft a Natural Language Database

Dorothee Beermann and Pavel Mihaylov

NTNU, Trondheim, Norway
and
Ontotext Lab, a company of Sirma Group, Sofia, Bulgaria

TypeCraft

With TypeCraft the user adds linguistic annotation to written material which is stored in a relational database from where it can be retrieved using multiple views.

Annotations are created on multiple tiers.

Texts as well as annotations are in Unicode. Annotated data can be exported to standard text editors (WORD, Open Office and LaTex) as well as to XML format. Archiving can be achieved through the export of XML with embedded DTD.

TC has been designed for projects on minority languages, joining linguists and native speakers. It is intuitive to use and allows wide and distributive usage.

TC uses PostgreSQL as database format and is written in the Java programming language. It can be freely used online with Mozilla Firefox, although at present a login is required.

TC runs on a server owned by NTNU.

We will show how textual data relevant for language documentation and/or for the presentation and publication of linguistic research can be annotated.

We demonstrate the basic functionality of TypeCraft including some of the basic data management tasks relevant for linguists.
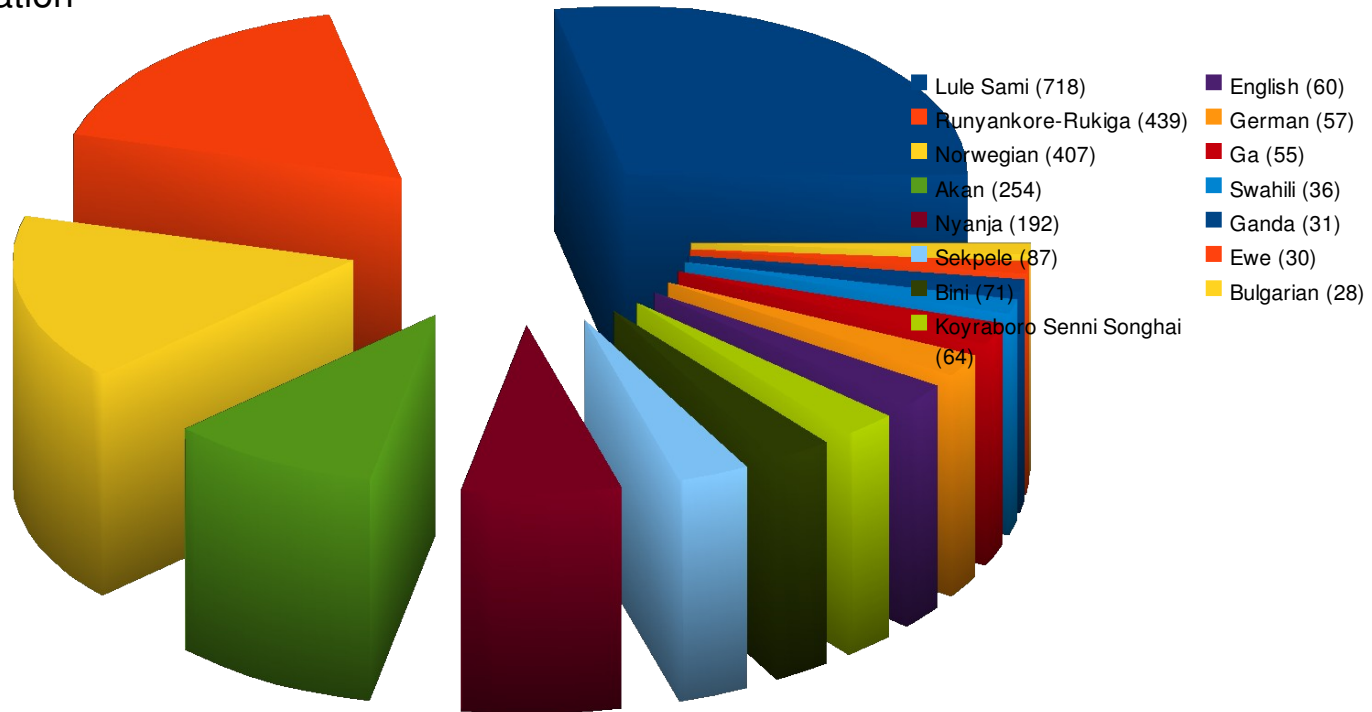
You will see the TypeCraft Editor and the TypeCraft Communication and Collaboration Tool.

You will learn which role TypeCraft can play in the management of linguistic resources.

NTNU
Innovation and Creativity

Some facts...

TypeCraft has at present
40 individual users, and
serves as collaboration
tool for 2 funded
research projects

Top 15 languages

(by number of phrases)

- Lule Sami (718)
- Runyankore-Rukiga (439)
- Norwegian (407)
- Akan (254)
- Nyanja (192)
- Sekpele (87)
- Bini (71)
- Koyraboro Senni Songhai (64)
- English (60)
- German (57)
- Ga (55)
- Swahili (36)
- Ganda (31)
- Ewe (30)
- Bulgarian (28)

NTNU
Innovation and Creativity

Review of the main functionality
Word level annotation (1)

Meta level and word level annotation

The TC-annotation tool

Meta tags – Choice of tag set

Collaboration or Privacy – Your Choice

The text-editor interface allows you to select domains.
A domain can be shared by several annotators.

Data can belong to a project so that it can be shared between
several people, or it may be private, e.g., when annotation is
done for graduate work or an academic publication.

Every user has a private domain, called MY TEXTS which
allows him or her to access private data (private and shared)from
the browser interface.

NTNU
Innovation and Creativity

SEARCH – select a language
Text Search

**search for title
or translation string**

    The TC-annotation tool

Search within phrases - input mask

The TC-annotation tool

Search within phrases - output

Cross-classification for phrase search

Search for construction types

Output of Search for phrases that have tense and aspect as annotation focus

The TypeCraft Publishing Tool
article preview

The TypeCraft collaboration tool project home page

The TypeCraft Research Networks

How do I start to use TypeCraft ?

Open Firefox   ww.mozilla.com/en-US/firefox

Go to   www.typecraft.org

Ask for a login
Login, go to *NEW TEXT*
Insert text that you would like to annotate
Enter the annotation editor - annotate
Next time you login you will find your newly annotated text in
MY TEXT

THAT SIMPLE

NTNU
Innovation and Creativity