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.
TypeCraft has at present 40 individual users, and serves as collaboration tool for 2 funded research projects.
Review of the main functionality

Word level annotation (1)
Meta level and word level annotation
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.
SEARCH – select a language
Text Search

TypeCraft
The Natural Language Database

Search for texts
Title string:
Title translation string:
Language: Any

Search
Reset

LanguageTitle
Bini Extraction in Edo
Bini Argument sharing in multi-verb constructions
Bini Argument sharing in multi-verb constructions
Bini Argument sharing in multi-verb constructions

Title translationContributorLast changed
Ota Ogie 2008-05-31 21:05
Ota Ogie 2008-05-18 18:05
Ota Ogie 2008-04-01 16:04
Ota Ogie 2008-04-01 16:04

search for title or translation string
Search within phrases - input mask

Cross-classification for phrase search
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

Locative prepositions in Runyakitara


Keywords: lexical semantics, spatial semantics, prepositional phrases, locative expressions, prepositional functions, attribute-value matrices, meaning spaces

Runyakitara provides a rich grammatical inventory for hosting spatial concepts. Next to three nominal class markers which encode deixis as well as topological notions, verbs may carry spatial suffixes. Also spatial nouns are use, some of which combine either with omu, corresponding to English in, or aha, which can mean on, at, near or around. Both omu and aha may occur in isolation as locative prepositions. Dependent on the grammatical context they can have several meanings, and this study gives an overview of the register of their uses, and the set of meanings that define their semantics. All examples used are contained in TypeCraft in a text named after this article (although when we created the text in TypeCraft we used Runyankore-Rukiga as language name, and put Runyakitara in brackets). To find this text by going to SEARCH FOR TEXTS in the wiki navigation bar on the left side of your screen, and search for Locative prepositions in Runyankore-Rukiga (Runyakitara).

Next to their locative use, illustrated below for omu with an example from TypeCraft

Tuarugaho tuahika Nakawa, omu burugwaizoba bwa Kampala.

1PL PRES leave IND LOC 1PL PRES reach Nakawa in SPTL west of Kampala
V V NPRO PREP N PREP NPRO

"We left that place and reached Nakawa, in the west of Kampala."

both prepositions have several other uses. They may for example express partitivity, as illustrated below with another example from the database

emiyembe emihango omuri yo.

emiyembe emihango omuri yo
IV CL4 mango DEF CL4 big PART 3PL
N ADJ PREP PRON

"the big ones of the mangos"

As seen in the example above omu and aha undergo morphological changes, a fact further discussed in the article.

For both prepositions the article provides an overview over possible readings. For omu for example we identified and defined the following readings:
The TypeCraft collaboration tool project home page

The MaLex Project

The MaLex project is a cooperation between the Centre for Language Studies (CLS) at the University of Malawi, Zomba, Malawi, and the Department of Language and Communication Studies (ISK) at the Norwegian University of Science and Technology (NTNU), Trondheim, Norway. The project is financed by NUFU. Prof. Lars Hallan, NTNU, is the Norwegian coordinator and Prof. Al Menje, University of Malawi, is the coordinator in Malawi.

The "Lex" in MaLex stands for "lexicon", and lexica will be among the results of the project. A second focal area of MaLex is an in-depth annotation of representative corpora of the languages of Malawi, starting with Chichewa.

The goals of TypeCraft within the MaLex project are:

- an in-depth linguistic annotation of representative text fragments from Chichewa, Giyawo and Citumbuka.

Within this overall goal, the annotation of technological terms used in the medical campaign against HIV and AIDS is of special importance.

Category: Funded Projects
The TypeCraft Research Networks

Verbconstructions cross-linguistically - Introduction

Lars Hellan
NTNU
June 2008

Presented here is an initiative for constructing an environment enabling the enumeration of verb constructions cross-linguistically. The aims are:

- For each language, that the enumeration be complete and transparent;
- Across languages, that the enumerations be comparable.

The environment includes a labeling system which, for any verb construction of a given language, provides a template for that construction type displaying its argument structure, in a fashion as transparent as possible. The template is constructed from a universally established inventory of labeling primitives.

The initiative has started with, on the one hand, a rather extensive inventory of Verb Constructions in Norwegian (based on the TROLL 1989 work, NorkompLeks 1996, and Norsource 2007), and on the other, a comparative survey of closely related languages of the Volta Basin Area (supported by The Legon-Trondheim project on Ghanaian languages).

The characterization of a construction type relates to at least the following parameters, when applicable:

- diathesis/argument operations (such as passive, causativisation, applicative formation),
- syntactic valency,
- semantic participants and situation type,
- particular patterns of agreement, including coreference patterns (such as 'equi' and 'raising' patterns, argument sharing, secondary predication), tense/aspect agreement, and subject and object marking.
How do I start to use TypeCraft?

Open Firefox

Go to TypeCraft

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