



# 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.

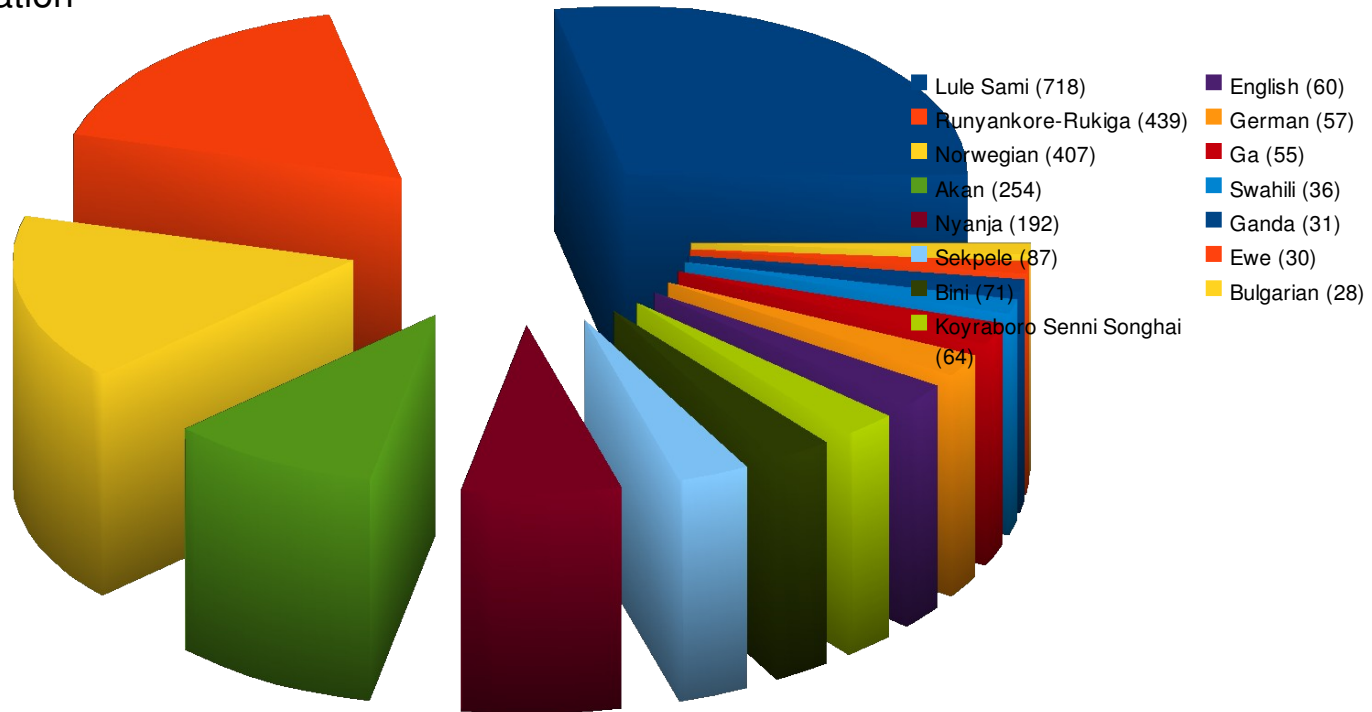
You will see the **TypeCraft Editor** and the **TypeCraft Communication and Collaboration Tool**.

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

# 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)



# Review of the main functionality

## Word level annotation (1)

Text editor - TypeCraft - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://ginnungagap.hf.ntnu.no/tc2wiki/Special:TypeCraft/Editor/11

Getting Started Latest Headlines

FoxLingo Webpage Text Services AutoTrans Search

### The Natural Language Database

special

## Text editor

Text Phrases

Text aha ishometro abaana nibahandii...

Save

Phrase: aha ishometro abaana nibahandiikiraho ebaruha

Free translation: at school children are writing letters

Annotation focus: nominal inflection

Construction parameters: Change vocative-right\_reordering--instrumentOBL-locativeModification-Fragmentation\_scenario-

Construction description: Locative inversion

<b>Latinised:</b>	àhà	ìshòmèrò	àbààná	nìbàhàndííkiráhò	èbàrúhà								
<b>Morpheme:</b>	aha	i	shomero	a	ba	ana	ni	ba	handiikir	a	ho	e	baruha
<b>Meaning:</b>	at		school			child			write				letter
<b>Gloss:</b>		CL5		I		CL2	PRS	CL2		APPL	IND	LOC	IV
<b>POS:</b>	PR	N		ILL		IMP		IMPERF					N

Word-level

Privacy policy About TypeCraft Disclaimers

Powered By MediaWiki

Done

start TCpresentations Control Panel My Documents Microsoft PowerPoint ... Text editor - TypeCra... 11:58

# Meta level and word level annotation

The screenshot shows the TypeCraft web application interface. The main content area displays the text "Die Wanderung beginnt in Todtnau..." with various annotation fields:

- Phrase:** Die Wanderung beginnt in Todtnau.
- Free translation:** The hike starts in Todtnau
- Annotation focus:** lexical semantics
- Construction:** ---locativeModification--
- parameters:**
- Construction description:**

Below these fields is a table showing the morphological and syntactic analysis of the words:

<b>Latinised:</b>	die	Wanderung	beginnt	in	Todtnau
<b>Morpheme:</b>	die	Wander ung	beginn t	in	Todtnau
<b>Meaning:</b>		<i>hike</i>	<i>start</i>		
<b>Gloss:</b>	DEF.SG.FEM.NOM		NMLZ.INCEP.PRES	STARTPNT	
<b>POS:</b>	DET	N	V	PREP	NPRO

Two grey arrows point to the right from the right side of the interface, indicating the levels of annotation:

- Construction-level:** Points to the "Construction" and "parameters" fields.
- Word-level:** Points to the table below.



## 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

Search for texts - TypeCraft - Mozilla Firefox

http://ginnungagap.hf.ntnu.no/tc2wiki/Special:TypeCraft/SearchText/

Getting Started Latest Headlines

FoxLingo Webpage Text Services AutoTrans Search

Dorothee my talk my preferences my watchlist my contributions log out

# TypeCraft

The Natural Language Database

typecraft

- New text
- My texts
- Search for texts
- Search for phrases
- POS tags
- Gloss tags

typecraftwiki navigation

- Main Page
- Community portal
- research
- current events
- Recent changes
- Random page
- Help
- Donations

search

Go Search

toolbox

- Upload file
- Special pages

special

## Search for texts

Title string:

Title translation string:

Language:

Search Reset

**search for title  
or translation string**

Language	Title	Title translation	Contributor	Last changed
Bini	<a href="#">Extraction in Edo</a>		Ota Ogie	2008-05-31 21:05
Bini	<a href="#">_Argument sharing in multi-verb constructions</a>		Ota Ogie	2008-05-18 18:05
Bini	<a href="#">_Argument sharing in multi-verb constructions</a>		Ota Ogie	2008-04-01 16:04
Bini	<a href="#">_Argument sharing in multi-verb constructions</a>		Ota Ogie	2008-04-01 16:04

Done

start TCpresentations Control Panel Microsoft Power... Search for texts... Microsoft Excel ... Inbox for dorot... Telefonnummer ... 13:01

# Search within phrases - input mask

Cross-classification  
for phrase search

# Search within phrases - output

Search for phrases - TypeCraft - Mozilla Firefox

http://ginnungagap.hf.ntnu.no/tc2wiki/Special:TypeCraft/SearchPhrase/

Getting Started Latest Headlines

Search for phrases - TypeCraft

Dorothee my talk my preferences my watchlist my contributions log out

# TypeCraft

The Natural Language Database

typecraft

- New text
- My texts
- Search for texts
- Search for phrases
- POS tags
- Gloss tags

typecraftwiki navigation

- Main Page
- Community portal
- research
- current events
- Recent changes
- Random page
- Help
- Donations

search

Go Search

toolbox

- Upload file
- Special pages

special

## Search for phrases

Search result (2 phrases found):

<input type="checkbox"/> LanguagePhrase	Translation	Contributor	Last changed
<input type="checkbox"/> Sekpele <a href="#">ɔ̃lɛ budi</a>	s/he is eating	cephas delalorm	2008-05-07 16:05
<input type="checkbox"/> Sekpele <a href="#">Kofi to budi ɔmɛ kpe sikpi</a>	Kofi's eating is frightening	cephas delalorm	2008-05-07 16:05

Export in HTML Export in LaTeX Export in XML

Prepare for wiki

Cross-classification for phrase search

# Search for construction types

Search for phrases - TypeCraft - Mozilla Firefox

http://ginnungagap.hf.ntnu.no/tc2wiki/Special:TypeCraft/SearchPhrase/

## TypeCraft

The Natural Language Database

special

### Search for phrases

Search result (22 phrases found):

Language	Phrase	Translation	Contributor	Last changed
<input type="checkbox"/> Runyankore-Rukiga	<a href="#">Akabaziirira aha kiteteevi ky'omuhara ky'okugaitwa.</a>	She sewed on the wedding dress for her daughter	Allen	2008-04-15 12:04
<input type="checkbox"/> Runyankore-Rukiga	<a href="#">Akasiiga aha kishushani kye ekisya.</a>	He painted on his new picture	Allen	2008-04-15 12:04
<input type="checkbox"/> Norwegian	<a href="#">Da han skulle til å gå, lo jeg av ham.</a>	As he was about to leave, I laughed at his expence.	Test	2008-03-01 16:03
<input type="checkbox"/> Norwegian	<a href="#">Mannen var i ferd med å bygge et stillas.</a>	The man was (in the process of) building/about to build a scaffold.	Test	2008-03-01 16:03
<input type="checkbox"/> Norwegian	<a href="#">Arne drev og løpte intervaller.</a>	Arne was running intervals.	Test	2008-03-01 16:03
<input type="checkbox"/> Norwegian	<a href="#">Karsten driver og jobber i drivhuset.</a>	Carsten is working in the greenhouse.	Test	2008-03-01 16:03
<input type="checkbox"/> Norwegian	<a href="#">Kari holder på med å gjøre lekser.</a>	Kari is doing her homework.	Test	2008-03-01 16:03
<input type="checkbox"/> Norwegian	<a href="#">Arne holder på å vokse.</a>	Arne is growing.	Test	2008-03-01 15:03
<input type="checkbox"/> Norwegian	<a href="#">Det holdt opp å regne.</a>	It stopped raining.	Test	2008-03-01 15:03
<input type="checkbox"/> Norwegian Bokmål	<a href="#">Det sitter en mann på kontoret og skriver dikt</a>	A man is writing poems in the office	Pavel	2008-01-19 20:01
<input type="checkbox"/> Swahili (generic)	<a href="#">Wanafunzi husoma vitabu.</a>	Students use to read books	Pavel	2008-01-19 20:01
<input type="checkbox"/> Logba	<a href="#">Ebitsie ofo memgba nu</a>	The child washed the plate	Pavel	2008-01-10 19:01
<input type="checkbox"/> Ga'anda	<a href="#">omucungwa kuqanagisibwa omwana</a>	An orange was made to be thrown by a child	Pavel	2007-09-10 01:09
<input type="checkbox"/> Runyankore-Rukiga	<a href="#">Kakuru yaaturisa omupiira</a>	Kakuru has made the ball to burst	Pavel	2007-08-01 17:08
<input type="checkbox"/> Runyankore-Rukiga	<a href="#">Abagvenyi bakashangwa omu nju</a>	Visitors were found in the house	Allen	2007-08-01 17:08
<input type="checkbox"/> Akan	<a href="#">kwaku nkutu ho</a>	Kwaku does not kneel there	Pavel	2007-08-01 17:08
<input type="checkbox"/> Ewe	<a href="#">devi sueawo yi suku</a>	The little child has gone to school	Pavel	2007-08-01 17:08
<input type="checkbox"/> Ganda	<a href="#">Abawala balungi</a>	Beautiful girls	Pavel	2007-08-01 17:08
<input type="checkbox"/> Sekpele	<a href="#">kofi kon</a>	Kofi really	Pavel	2007-08-01 17:08
<input type="checkbox"/> Runyankore-Rukiga	<a href="#">omu nju omu haataahamu abaana</a>	in this house children have entered	Allen	2007-08-01 17:08
<input type="checkbox"/> Runyankore-Rukiga	<a href="#">abaana nibahandiikira ebaruha aha ishomeru</a>	Children are writing letters at school	Allen	2007-08-01 17:08
<input type="checkbox"/> Runyankore-Rukiga	<a href="#">Ebitabo biteho aha meeza</a>	Put the books on the table	Allen	2007-08-01 17:08

Export in HTML    Export in LaTeX    Export in XML

Prepare for wiki

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

# The TypeCraft Publishing Tool article preview

Locative prepositions in Runyakitara - TypeCraft - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://ginnungagap.hf.ntnu.no/tc2wiki/Locative\_prepositions\_in\_Runyakitara

Getting Started Latest Headlines

FoxLingo Webpage Text Services AutoTrans Search

typecraft article discussion edit history protect delete move watch

- New text
- My texts
- Search for texts
- Search for phrases
- POS tags
- Gloss tags

typecraftwiki navigation

- Main Page
- Community portal
- research
- current events
- Recent changes
- Random page
- Help
- Donations

search

Go Search

toolbox

- What links here
- Related changes
- Upload file
- Special pages
- Printable version
- Permanent link

## Locative prepositions in Runyakitara

**Beermann, D. and Allen Asiimwe (2008) Locative prepositions in Runyankitara - A study in Bantu** [edit]

**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 burugwaizooba bwa Kampala.**

Tuarugaho	tuahika	Nakawa omu	burugwaizooba	bwa	Kampala
tu a	rug a	ho tu a	hika nakawa omu	burugwa izooba	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
e mi yembe e	mi hango	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:



Done

start TCpresentations Control Panel Microsoft Power... Locative preposi... Microsoft Excel ... Inbox for doroth... Telefonnummer t... 13:36

# The TypeCraft collaboration tool project home page

The MaLex Project - TypeCraft - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://ginnungagap.hf.ntnu.no/tc2wiki/The\_MaLex\_Project

Getting Started Latest Headlines

FoxLingo Webpage Text Services AutoTrans Search

Dorothee my talk my preference my watchlist my contributions log out

My preferences

# TypeCraft

The Natural Language Database

article discussion edit history protect delete move watch

## The MaLex Project



NTNU-Dravoll, Trondheim Norway

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 Hellan, NTNU, is the Norwegian coordinator and Prof. Al Mtenje, University of Malawi, is the coordinator in Malawi.



University of 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, Ciyawo 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

http://ginnungagap.hf.ntnu.no/tc2wiki/Special:Preferences

start TCpresentations Control Panel Microsoft Power... The MaLex Proje... Microsoft Excel ... Inbox for dorothe... Telefonnummer t... 13:40

Public Window for Funded Projects

# The TypeCraft Research Networks

Verbconstructions cross-linguistically - Introduction - TypeCraft - Mozilla Firefox

http://ginnungagap.hf.ntnu.no/tc2wiki/Verbconstructions\_cross-linguistically\_-\_Introduction

Getting Started Latest Headlines

FoxLingo Webpage Text Services AutoTrans Search

Dorothee my talk my preferences my watchlist my contributions log out

## TypeCraft

The Natural Language Database

article discussion edit history protect delete move watch

### 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,

Done

start TCpresentations Control Panel Microsoft Power... Verbconstruction... Microsoft Excel ... Inbox for dorothe... Telefontnummer t... 13:41

# How do I start to use TypeCraft ?



Open Firefox



[www.mozilla.com/en-US/firefox](http://www.mozilla.com/en-US/firefox)

Go to



[www.typecraft.org](http://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

