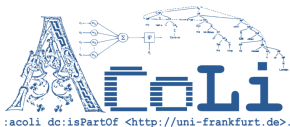# Ligt: A vocabulary for representing interlinearly glossed text

Max Ionov

ionov@cs.uni-frankfurt.de

Applied Computational Linguistics Lab, Goethe University Frankfurt, Germany

LD4LT Workshop, 04.09.21

:acoli dc:isPartOf <http://uni-frankfurt.de>.

[yʉg]

LiODi
er R
Linked Open
Dictionaries

GOETHE
UNIVERSITÄT
FRANKFURT AM MAIN

Outline

Interlinear glossed texts (IGT)

- Notation widely used in linguistic research

Interlinear glossed texts (IGT)

- Notation widely used in linguistic research
  - Language documentation
  - Demonstration of language phenomena

Interlinear glossed texts (IGT)

- Notation widely used in linguistic research
  - Language documentation
  - Demonstration of language phenomena
- Alignment between elements on different lines

Interlinear glossed texts (IGT)

- Notation widely used in linguistic research
  - Language documentation
  - Demonstration of language phenomena

- Alignment between elements on different lines

(1)   Kajirringku im katim jawurt kengkaruyu.

    Kajirri-ngku im   kat-im jawurt kengkaru-yu.
    woman-ERG 3SG cut-TR tail   kangaroo-DAT

    'The old woman chopped off the kangaroo's tail.'
    (Gurindji Kriol)

IGT and interoperability

1. A lot of ways to create and store IGT

IGT and interoperability

1. A lot of ways to create and store IGT:
   - Word
   - PDF
   - XML
   - FLEx/Toolbox/xigt

IGT and interoperability

1. A lot of ways to create and store IGT:
   - Word
   - PDF
   - XML
   - FLEx/Toolbox/xigt tools and formats for language documentation

IGT and interoperability

1. A lot of ways to create and store IGT:
   - Word
   - PDF
   - XML
   - FLEx/Toolbox/xigt tools and formats for language documentation

2. Grammatical categories can vary
   - Inventories tailored for a language
   - Different ways of writing down the same thing: COMP / COMPL / compl
   ⇒ Inconsistencies between sources and authors

IGT and interoperability

1. A lot of ways to create and store IGT:
   - Word
   - PDF
   - XML
   - FLEx/Toolbox/xigt tools and formats for language documentation

2. Grammatical categories can vary
   - Inventories tailored for a language
   - Different ways of writing down the same thing: COMP / COMPL / compl
   ⇒ Inconsistencies between sources and authors *usually* documented

IGT and interoperability

1. A lot of ways to create and store IGT:
   - Word
   - PDF
   - XML
   - FLEx/Toolbox/xigt tools and formats for language documentation

2. Grammatical categories can vary
   - Inventories tailored for a language
   - Different ways of writing down the same thing: COMP / COMPL / compl
   - ⇒ Inconsistencies between sources and authors *usually* documented
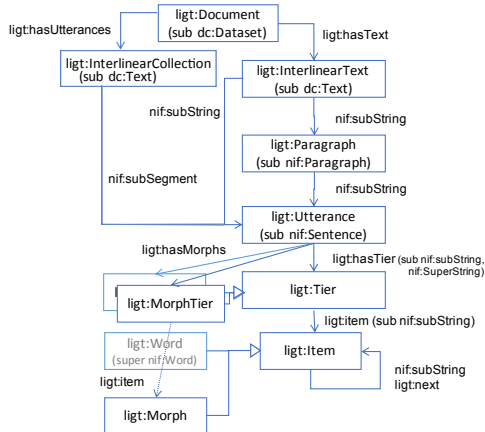
⇒ **Ligt**: Linked Data representation for IGT

Outline

Ligt model

Outline

1. Why

2. What

3. **Where**

Version history

- 2017: first version

Version history

- 2017: first version
- 2019: polished and presented at LDK; applied at SD-LLOD

Version history

- 2017: first version
- 2019: polished and presented at LDK; applied at SD-LLOD
- 2021: extended for non-coherent texts and presented at LDK

Fieldwork tools

Converters from popular fieldwork and language documentation
tools

Fieldwork tools

Converters from popular fieldwork and language documentation tools:

- Toolbox
- FLEx
- Xigt format

https://github.com/acoli-repo/LLODifier

ELAR

Nordhoff 2020: Ligt applied to 280 endangered language resources
(presented @ LAW-2020)

APiCS

Ionov 2021: Ligt applied to the Atlas of Pidgin and Creole
Languages (presented @ LDK-2021, yesterday)

Next steps

- User-friendly interfaces
  - Converters
  - Readers
  - Search

Next steps

- User-friendly interfaces
  - Converters
  - Readers
  - Search

- Integration with CoNLL-RDF

Next steps

- User-friendly interfaces
  - Converters
  - Readers
  - Search

- Integration with CoNLL-RDF $\Rightarrow$ integrating sub-word information into corpora

Next steps

- User-friendly interfaces
  - Converters
  - Readers
  - Search

- Integration with CoNLL-RDF $\Rightarrow$ integrating sub-word information into corpora

- Integration with `cqp4rdf`

Next steps

- User-friendly interfaces
  - Converters
  - Readers
  - Search

- Integration with CoNLL-RDF $\Rightarrow$ integrating sub-word information into corpora

- Integration with cqp4rdf $\Rightarrow$ search query language and interface

Thank you!
Any questions?

`https://github.com/acoli-repo/ligt`