RELISH (Rendering Endangered Languages Lexicons Interoperable through Standards Harmonization)

Justin Petro

The LINGUIST List
Overview

• Differences in Digital Standards for Lexicon Creation
  • U.S. (LEGO Project)
    • GOLD
    • LIFT/LEGO
  • Europe (LEXUS)
    • ISOcat
    • LMF

• Scope and Goals of RELISH Project
  • How RELISH harmonizes these standards
  • Rendering six lexicons into RELISH format (as proof of concept)
  • Ultimate goal of the project
Part 1: Diverging Standards

GOLD, ISOcat, LIFT, LMF
• **General Ontology for Linguistic Description**
• Formalized account of generally agreed upon linguistic concepts – “codifies general knowledge of the field”
• As an ontology, it allows complex relationships among concepts, although the only relationship currently instantiated is the hierarchical one of parent-child
• Can be used to promote interoperability and searching amongst disparate data sets (as seen in LEGO presentation)
• **GOLD Website**
Vegetable Gender (Concept)

http://purl.org/linguistics/gold/VegetableGender

Definition:
Vegetable gender refers to inanimates and exists in some four-way gender systems, e.g., masculine, feminine, neuter, and vegetable as in BiniŋGun-wok [Evens 2003: 202].

Usage Notes

Examples

Blackfoot example of plural number and animate gender.
2009-06-04 13:28:09

man- ne man- me
III- food VEG- good
good food

References:
Mithun (1999-99) FROM Frantz (1991:8-9)

Properties/Values/Definition

User Submitted Issues

Bininj Gunwok example for vegetable gender - Dunstan Brown
Status: Pending
2009-09-16 09:06:50

Blackfoot example of plural number and animate gender.
Usage Notes

Examples

Blackfoot example of plural number and animate gender.

Language Code: bla

man- me man- mak
III- food VEG- good
good food

References:

Properties/Values/Definition

User Submitted Issues

Bininj Gunwok example for vegetable gender - Dunstan Brown

Status: Pending

Blackfoot example of plural number and animate gender.

man- me man- mak
III- food VEG- good
good food

References:

Comment: the example is from Bininj Gunwok but the references refer to other languages.

References:
ISO Data Category Registry (ISOcat)

- Developed by Max Planck Institute for Psycholinguistics, (Nijmegen, The Netherlands.)
- Terminology Registry for Linguistic Concepts, currently without hierarchy (work is being done to change this)
- Maps divergent data to a common term set: Specific terms can be associated with data element names using a namespace reference and a datcat PID
- Used in LMF-compliant digital lexicons
Digital Lexicon Standards

- **LIFT (Lexicon Interchange Format)**
  - Developed by SIL as a digital interchange format between lexicons created using different software tools
  - Schema restricted XML format
  - LL-LIFT: further restricted version of LIFT used by LEGO project
- **LMF (Lexical Markup Framework)**
  - ISO/24613:2008 standard for natural language processing (NLP) and machine-readable dictionary (MRD) lexicons.
  - Template for multilingual lexicon schemas
  - Used by LEXUS
Part 2: The RELISH Project

Its Scope and Ultimate Goals
**Scope of RELISH Project: Integration of Digital Standards**

- Integrate GOLD Concepts into ISOcat Registry
  - **GOLD in ISOcat**
- Create interchange format between lexicons output by LEXUS and those by LEGO
  - LMF serialization based on LIFT
  - LEGO and LEXUS will make RELISH output of lexicons available
<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Administration Status</th>
<th>Registration Status</th>
<th>Check</th>
<th>Type</th>
<th>Owned by</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>TransitiveVerb</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb1</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb2</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb3</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb4</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb5</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb6</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb7</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb8</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb9</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
<tr>
<td>TransitiveVerb10</td>
<td>1.0</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
</tbody>
</table>

**VegetableGender**

1.1 Administration Record

- **Identifier**: VegetableGender
- **Version**: 1.10
- **Registration Status**: private
- **Status**:
  - Administration: private
  - Status: private
- **Justification**: The GOLD ontology was developed as part of the E-MELD project funded by the NSF (BCS0729644 and BC05094504). The concepts were reviewed by 50 noted typologists and documentary linguists at the E-MELD 2005 workshop (http://emeld.org/workshop/2005/proceeding.html). They have since been refined by the GOLD Community (http://linguistics-ontology.org).
- **Origin**: GOLD (General Ontology for Linguistic Description) for the latest version see http://linguistics-ontology.org

<table>
<thead>
<tr>
<th>Name</th>
<th>Version</th>
<th>Administration Status</th>
<th>Registration Status</th>
<th>Check</th>
<th>Type</th>
<th>Owned by</th>
<th>Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>VegetableGender</td>
<td>1.10</td>
<td>private</td>
<td>private</td>
<td>✔</td>
<td>simple</td>
<td>gold-user</td>
<td>public</td>
</tr>
</tbody>
</table>

1.1.1 Creation

- **Creation Date**: 2010-08-08

1.1.2 Last Change

- **Last Change Date**: 2010-07-16
Grant Project Goal: Harmonize Standards
Proof of Concept for this goal: Take 6 Lexicons and convert them into RELISH format
    - From LEGO: Mocovi, Western Sisaala, and Fulfulde
    - From LEXUS: Udi, Wichita, and Tuva
    - Demonstrates utility of RELISH format
Ultimate Goal of Project:
    - Provide widest accessibility to lexicons of endangered languages
    - Make lexical data interoperable for not only linguists, but historians, anthropologists, geneticists, and other scientists interested in language
    - Participate in the trend to encourage data interoperability and integrated cyber-infrastructure.